

VU Research Portal

De taal van 'Den Haag'

Renkema, E.J.

1981

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Renkema, E. J. (1981). *De taal van 'Den Haag': Een kwantitatief -stilistisch onderzoek naar aanleiding van oordelen over taalgebruik*. [, Vrije Universiteit Amsterdam].

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

J. Renkema

De taal van ‘Den Haag’

Een kwantitatief-stilistisch
onderzoek
naar aanleiding van
oordelen over taalgebruik

NK
11

Staatsuitgeverij

De taal van 'Den Haag'

Een kwantitatief-stilistisch onderzoek
naar aanleiding van
oordelen over taalgebruik

door Jan Renkema

Eveneens verschenen als proefschrift ter verkrijging van de graad van
doctor in de letteren aan de Vrije Universiteit te Amsterdam.

Promotor : Prof.dr. B. Siertsema

Copromotor: Prof.dr. H. Brandt Corstius

© J. Renkema

No part of this book may be reproduced in any form, by print,
photoprint, microfilm, or any other means without written permission
of the copyright owner.

Niets uit deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt
door middel van druk, fotocopie, microfilm of op welke wijze ook, zonder
voorafgaande schriftelijke toestemming van de houder van het copyright.

ISBN 90 12 03487 6

DANKWOORD

Bij het verschijnen van deze proeve van bekwaamheid in wetenschappelijk onderzoek past een woord van dank aan allen die mij hebben geholpen.

Mijn promotor deed meer dan duistere probleemstellingen verhelderen: haar enthousiasme hielp mij over dode punten heen. Mijn copromotor behoeftte mij voor een te grootse, en daardoor weinig vruchtbare, opzet en maakte voor mij het onderzoek per computer mogelijk.

Een belangrijke stimulans voor dit onderzoek ging uit van de hoofd-directeur van het Staatsdrukkerij- en Uitgeverijbedrijf, Th.H. Oltheten, en van wijlen de heer A. Vondeling, die mij in mijn vorige werkkring alle ruimte gaven voor dit onderzoek naar verschillen in taalgebruik.

Zonder hulp van de programmeurs A.C. Veldkamp (Mathematisch Centrum Amsterdam), J.J.H.M. van der Ven (Rekencentrum TH-Eindhoven), G. van der Steen en J. Skolnik (Computerafdeling Faculteit der Letteren, Universiteit van Amsterdam) en de statistici R.D. Gill en R.J.M.M. Does (Mathematisch Centrum Amsterdam) had ik aanzienlijk meer problemen gehad bij het toetsen van hypothesen over taalkundige verschijnselen in overheids teksten.

Dank ben ik ook verschuldigd aan degenen die kritisch commentaar hebben geleverd op gedeelten van deze studie, leden van: de VU-linguïstenkring, de Werkgemeenschap Sociolinguïstiek (werkgroep taalvariatie en taalverandering) van de Stichting Taalwetenschap, de werkgroep begripelijkheidsonderzoek van de Stichting Onderzoek Massacommunicatie en de Vereniging Interuniversitair Overleg Taalbeheersing, en aan A.J. van Oosten en P.C. Uit den Boogaart.

Het onderzoek zou aanzienlijk langer hebben geduurd zonder hulp van de codeerassistenten J. van Beusekom en M. Cox (stenografen bij de Tweede Kamer), en mevrouw H.M.J. van der Laan-Nieuwenhuizen, die ook het tikwerk voor haar rekening nam.

J. Renkema

INHOUD

1	INLEIDING	1
1.1	<i>Achtergrond</i>	1
1.2	<i>Probleemstelling</i>	2
1.3	<i>Hoofdstukindeling</i>	3
2	STILISTIEK, THEORIE EN METHODE	6
2.1	<i>Inleiding</i>	6
2.2	<i>Stijlopvattingen</i>	8
2.3	<i>Stijl als afwijking van een norm</i>	11
2.3.1	De norm?	11
2.3.2	Bezwaren tegen de opvatting van stijl als afwijking van een norm	13
2.4	<i>De kwantitatieve methode</i>	16
2.4.1	Polariserende opvattingen	16
2.4.2	Bezwaren tegen de kwantitatieve methode	19
2.4.3	Algemene opmerkingen over de kwantitatieve methode	22
3	OORDELEN OVER DE TAAL VAN 'DEN HAAG'	24
3.1	<i>Inleiding</i>	24
3.1.1	Beperkingen	24
3.1.2	De opbouw van dit hoofdstuk	27
3.2	<i>De aard van de oordelen</i>	28
3.3	<i>Algemene oordelen</i>	30
3.3.1	Zes publikaties	30
3.3.2	Samenvatting	32
3.4	<i>Oordelen over zinsbouw en woordkeus</i>	33
3.4.1	Onpersoonlijk	33
3.4.1.1	Naamwoordstijl	33
3.4.1.2	Lijdende vorm	35
3.4.1.3	Het menselijk element	36
3.4.2	Ingewikkeld	37
3.4.2.1	Bijzinnen en tussenzinnen	37
3.4.2.2	Tangconstructies	38
3.4.2.3	Andere zinsbouwverschijnselen	39
3.4.2.4	Moeilijke woorden	40
3.4.3	Wijdlopig	40
3.4.4	Traditioneel	41
3.4.4.1	Boekentaal	42
3.4.4.2	Clichématig taalgebruik	43
3.5	<i>De waarde van de oordelen</i>	44
4	WOORDFREQUENTIES	47
4.1	<i>Inleiding</i>	47
4.2	<i>Het codeersysteem C_3C</i>	48
4.2.1	Overzicht C_3C	48
4.2.2	Toelichting	50
4.3	<i>Enkele opmerkingen bij het codeersysteem C_3C</i>	68
4.3.1	Nomenclatuur	69

4.3.2	Onderscheidingen	70
4.3.3	Slordigheden	73
4.4	<i>De Frequentielijsten</i>	75
4.5	<i>Enkele opmerkingen bij de frequentielijsten</i>	76
5	DE STEEKPROEF, SELECTIE EN BEWERKING	79
5.1	<i>Inleiding</i>	79
5.1.1	Belang van de verslaggeving	79
5.1.2	Het bronnenmateriaal	81
5.1.3	Bezwaren tegen het onderzoeksmateriaal	82
5.1.4	De omvang van de steekproef	86
5.2	<i>De steekproef</i>	87
5.2.1	Drie arbitraire beslissingen	87
5.2.2	Het selecteren van de pagina's	88
5.2.3	Het selecteren van de fragmenten	90
5.3	<i>Herkomst van het onderzoeksmateriaal met voorbeelden</i>	92
5.4	<i>Tekstbewerking</i>	97
6	HYPOTHESES EN TOETSING	102
6.1	<i>Inleiding</i>	102
6.1.1	Oordelen buiten het C ₃ C-kader	102
6.2	<i>Hypotheses</i>	103
6.2.1	Onpersoonlijk	104
6.2.1.1	Naamwoordstijl	104
6.2.1.2	Lijdende vorm	108
6.2.1.3	Het menselijk element	112
6.2.2	Ingewikkeld	118
6.2.2.1	Bijzinnen en tussenzinnen	118
6.2.2.2	Tangconstructies	120
6.2.2.3	Andere zinsbouwverschijnselen	125
6.2.2.4	Moeilijke woorden	127
6.2.3	Wijdlopig	128
6.2.4	Traditioneel	129
6.3	<i>Toetsing</i>	133
6.3.1	Wijze van toetsing	133
6.3.2	Overzicht van de resultaten	137
7	DISCUSSIE	141
7.1	<i>Inleiding</i>	141
7.1.1	De operationalisering van de oordelen	141
7.1.2	De indeling van de hypotheses	143
7.1.3	Zinslengte	145
7.2	<i>Bespreking van de uitkomsten</i>	147
7.2.1	Onpersoonlijk	147
7.2.1.1	Naamwoordstijl	147
7.2.1.2	Lijdende vorm	149
7.2.1.3	Het menselijk element	152
7.2.2	Ingewikkeld	153
7.2.2.1	Bijzinnen en tussenzinnen	153
7.2.2.2	Tangconstructies	154
7.2.2.3	Andere zinsbouwverschijnselen	156
7.2.2.4	Moeilijke woorden	156
7.2.3	Wijdlopig	156
7.2.4	Traditioneel	157
7.4	<i>Conclusies</i>	157
7.5	<i>Verder onderzoek</i>	160

SAMENVATTING 164

SUMMARY 167

Bijlage 1 OVERZICHT CODEERSYSTEEM C_3C 170

Bijlage 2 CODES EN FREQUENTIES 173

Bijlage 3 UITKOMSTEN VAN DE TOETSINGEN 177

Bijlage 4 GRAFIEKEN ZINSLENGTE 220

LITERATUUR 222

REGISTER 232

1 INLEIDING

1.1 Achtergrond

De taal van 'Den Haag' is voorwerp van aanhoudende kritiek. Talrijk zijn de klachten over vaag, onpersoonlijk, omslachtig en ouderwets taalgebruik van ambtenaren. Het taalgebruik van de overheid wordt door velen gekenschetst als onnodig ingewikkeld. Of zoals *Orwell* (1968:133) het zegt: Als u de volgende zin uit het boek *Prediker* naar een overheidsinstelling stuurt:

"I returned, and saw under the sun, that the race is not to the swift, nor the battle to the strong, neither yet bread to the wise, nor yet riches to men of understanding, nor yet favour to men of skill; but time and chance happeneth to them all", dan is de kans groot dat diezelfde zin als volgt weer uit de ambtelijke molen komt:

"Objective consideration of contemporary phenomena compels the conclusion that success or failure in competitive activities exhibits no tendency to be commensurate with innate capacity, but that a considerable element of the unpredictable must invariably be taken into account." Zie voor Nederlandstalige parodieën bijvoorbeeld *Van Vierssen Trip* (1969) en het 'Officieel schrijven' in *Kousbroek* (1978:49,50).

Naar aanleiding van de klachten over de taal van 'Den Haag' zijn ook ambtelijke commissies in het leven geroepen. De meeste bekendheid kreeg de Commissie Duidelijke Taal. Deze commissie, bestaande uit vertegenwoordigers van departementen en enkele andere overheidsinstanties, de parlementaire pers en enkele neerlandici, is in 1973 ingesteld door de toenmalige voorzitter van de Tweede Kamer, dr. A. Vondeling. De commissie probeert een antwoord te geven op de vraag hoe het taalverkeer tussen overheid en burger verbeterd kan worden.

De kritiek op het taalgebruik van ambtenaren, juristen en politici dateert overigens niet van de laatste tijd. In 1826 schreef *Lulofs* een brochure over het taalgebruik in de nieuwe wetboeken, waarin hij vertalingen voor 'stadhuiswoorden' voorstelde. In 1844 deed *Van Lennep* een voorstel om alle uitheemse woorden uit de Grondwet te verwijderen, en in het algemeen de tekst van de Grondwet te verduidelijken - een voorstel waarop overigens veel kritiek kwam van *Thorbecke*. Zie *Van Assen* (1844) en *De Vries* (1948:13,14).

Naar aanleiding van alle - overigens te pas en te onpas gegeven - kritiek op de taal van 'Den Haag' zijn er pogingen gedaan om overheids-teksten te herschrijven (zie o.a. *Vondeling en Renkema* 1976) en is er onderzoek verricht naar de leesbaarheid van overheidsteksten (zie o.a. *Korsten* 1976 en *Renkema* 1976). Maar de resultaten van herschrijvingen en leesbaarheidsonderzoek bleven bijzonder mager. Tegen de hier geschetste achtergrond moet de probleemstelling in de volgende paragraaf gezien worden.

1.2 Probleemstelling

Gezien de veelal negatieve oordelen over de taal van 'Den Haag', gegeven de mogelijkheid dat ambtenarentaal te parodiëren valt, en gelet op de pogingen die gedaan zijn om publikaties van de overheid te herschrijven, lag de volgende vraag voor de hand: Wat is er zo opvallend aan het taalgebruik in teksten van de overheid? Uiteraard kan deze vraag ook anders geformuleerd worden. Bijvoorbeeld: Onderscheidt ambtelijke stijl zich van andere stijlen, en zo ja, waarin? Of: Wat is het verschil tussen overheidstaal en meer alledaags taalgebruik of andere typen taalgebruik?

Deze vragen gaan vooraf aan allerlei andere vragen, zoals: Waarom verschilt het taalgebruik van de overheid van andere vormen van taalgebruik? Of: Is het te bewijzen dat teksten van de overheid moeilijker leesbaar zijn dan andere teksten? Of: Hoe komt het dat veel lezers die en die oordelen hebben over het taalgebruik van ambtenaren? De vraag naar het verschil tussen ambtelijke stijl en andere stijlen moet dus eerst beantwoord worden voordat er verder onderzoek kan plaatsvinden.

Het probleem dat in dit onderzoek centraal staat, kan dan ook als volgt worden geformuleerd: Onderscheidt het taalgebruik in teksten van de overheid zich van het taalgebruik in andere teksten? Zo ja, waarin? De aldus geformuleerde probleemstelling gaat uit van de vooronderstelling dat er verschillende stijlen of, zo men wil, verschillende typen taalgebruik of variaties in taalgebruik bestaan. Deze vooronderstelling houdt in dat er antwoord gegeven kan worden op de volgende twee vragen: 1. Wat is stijl? 2. Hoe kan aangetoond worden dat er verschillen bestaan tussen typen taalgebruik?

De vooronderstelling dat er verschillende stijlen of typen taalgebruik bestaan, zal in deze studie niet ter discussie staan. Wel zullen de problemen worden behandeld die ontstaan, wanneer we een antwoord willen formuleren op de twee vragen, die door deze vooronderstelling worden opgeroepen.

Maar het centrale probleem leidt ons niet alleen naar deze twee vragen.

Andere vragen zijn:

1. Op welke verschijnselen moet men letten als men wil onderzoeken of een bepaald type taalgebruik zich onderscheidt van een ander type taalgebruik?
2. Welk beschrijvingsmodel moet worden gehanteerd om eventuele verschillen in kaart te brengen?
3. Welke teksten kunnen dienen om hypothesen over verschillen te toetsen?
4. Welke hypothesen worden opgesteld over eventuele verschillen?
5. Hoe worden de hypothesen getoetst?

Ook deze vragen moeten beantwoord worden voor het eigenlijke onderzoek van start kan gaan. De hier geschetste probleemstelling leidde tot een onderzoeksopzet die in de volgende paragraaf verduidelijkt zal worden aan de hand van de hoofdstukindeling in deze studie.

1.3 Hoofdstukindeling

Zoals uit de vorige paragraaf blijkt, ligt mijn probleemstelling op het terrein van de stilistiek. Een direct gevolg hiervan is dat deze studie begint met een hoofdstuk over stijl. In deze studie worden overigens de begrippen 'stijl', 'type taalgebruik' en 'variatie in taalgebruik' door elkaar gebruikt. Hierbij moet dan vooral gedacht worden aan verschijnselen die vallen onder zinsbouw en woordsoortverdelingen of het gebruik van bepaalde categorieën woorden.

In hoofdstuk 2, *Stilistiek, theorie en methode*, wordt eerst een overzicht gegeven van de stand van zaken binnen de stilistiek, voorzover dat voor de taalkunde van belang is. Daarna verdedig ik een bepaalde stijltheorie, namelijk stijl als afwijking van een (verwante) norm. Als verwante norm voor de taal van 'Den Haag' wordt hier voorgesteld het taalgebruik in kranten en populair-wetenschappelijke publikaties. Vervolgens wordt beargumenteerd waarom ik bij het verzamelen van feiten de methode heb gevolgd van de kwantitatieve stilistiek en niet die van de kwalitatieve stilistiek.

Hoofdstuk 3, *Oordelen over de taal van 'Den Haag'*, behandelt de vraag: Op welke verschijnselen moet gelet worden, wanneer men wil onderzoeken of een type taalgebruik zich onderscheidt van een ander type taalgebruik? Het hier gegeven antwoord luidt als volgt: Probeer die verschijnselen in het onderzoek te betrekken, die genoemd worden in de oordelen over het te onderzoeken type taalgebruik. Ik beperk me hier dus tot intuïtieve oordelen over een bepaald type taalgebruik. In dit hoofdstuk wordt een

overzicht gegeven van de oordelen over het taalgebruik van de overheid, zoals die te vinden zijn in kranteartikelen, in publikaties over verzorgd taalgebruik, over de taal van de overheid enz. Met nadruk moet hier gesteld worden dat het normatieve karakter van deze oordelen buiten beschouwing blijft. Het gaat er - populair gezegd - niet om of ambtenaren goed of slecht schrijven. Hoofddoel van deze beschrijvende studie is: te onderzoeken of er, naar aanleiding van deze kwalitatieve oordelen, volgens de kwantitatieve methode verschillen zijn aan te wijzen tussen de taal van 'Den Haag' enerzijds en 'krantetaal' en 'populair-wetenschappelijke proza' anderzijds.

In hoofdstuk 4, *Woordfrequenties*, wordt het in deze studie gekozen beschrijvingsmodel voor het in kaart brengen van verschillen, toegelicht met talrijke voorbeelden en van commentaar voorzien. Het betreft hier het codeersysteem zoals dat is toegepast in de publikatie waaraan de titel van dit hoofdstuk is ontleend 'Woordfrequenties in geschreven en gesproken Nederlands' (red., P.C. Uit den Boogaart, 1975). Globaal gezegd komt het systeem erop neer, dat aan elk woord in een tekst bepaalde informatie over grammaticale functie en eventuele morfologische eigenschappen wordt toegevoegd met behulp van een drie- en soms viercijfersysteem. Met dit codeersysteem zijn al grote aantallen fragmenten gecodeerd uit vijf verschillende categorieën teksten, namelijk: dagbladen, opiniebladen, gezinsbladen, romans en novellen, en populair-wetenschappelijke boeken. Op basis hiervan zijn frequentielijsten gepubliceerd van o.a. woordvormen en woordsoorten. In deze studie is dankbaar gebruik gemaakt van de op deze wijze beschikbare gegevens over het taalgebruik in dagbladen, opiniebladen en populair-wetenschappelijke boeken.

In hoofdstuk 5, *De steekproef, selectie en bewerking*, wordt beschreven hoe tekstgedeelten - ik heb mij beperkt tot geschreven taal - zijn geselecteerd en bewerkt om oordelen over de taal van 'Den Haag' te onderzoeken. Het betreft hier gedeelten uit de correspondentie tussen regering en Staten-Generaal. In dit hoofdstuk zijn ook enkele voorbeelden opgenomen van tekstgedeelten die tot deze correspondentie behoren. Bij de selectie is een onderscheid gemaakt tussen teksten van de regering en teksten van de Staten-Generaal, opdat ook eventuele verschillen tussen deze teksten aan het licht zouden kunnen komen.

In hoofdstuk 6, *Hypotheses en toetsing*, volgen de hypotheses die naar aanleiding van de oordelen (hoofdstuk 3) binnen het kader van het codeersysteem (hoofdstuk 4) zijn opgesteld over kenmerken van de taal

van 'Den Haag'. Bij elke hypothese wordt aangegeven op welke punten de steekproef wordt vergeleken met de teksten uit 'Woordfrequenties' die als verwante norm worden beschouwd, namelijk de corpora 'dagbladen', 'opiniebladen' en 'populair-wetenschappelijke boeken'. Na een beschrijving van de wijze van toetsing volgt aan het einde van dit hoofdstuk een samenvattend overzicht van de resultaten.

In hoofdstuk 7, *Discussie*, worden de uitkomsten besproken, conclusies getrokken en mogelijkheden aangegeven voor verder onderzoek. Tot slot volgt nog een samenvatting.

2 STILISTIEK, THEORIE EN METHODE

2.1 Inleiding

In dit hoofdstuk komen aan de orde de stijlopvatting die in deze studie wordt gehanteerd (theorie) en de wijze waarop de stilistische feiten zijn verzameld (methode).

Over theorie gaan de paragrafen 2.2 en 2.3. Nu ja, theorie? Wanneer wij de definitie van *De Groot* (1968:42) nemen, dan moeten we ons afvragen of het woord theorie in de stilistiek wel gebruikt mag worden: "een systeem van logisch samenhangende, met name niet-strijdige, beweringen, opvattingen en begrippen betreffende een werkelijkheidsgebied, die zo zijn geformuleerd, dat het mogelijk is er toetsbare hypothesen uit af te leiden."

Het woord 'theorie' moet binnen de stilistiek eerder letterlijk worden opgevat: zienswijze, beschouwingswijze, opvatting. In 2.2 wordt een overzicht gegeven van de verschillende opvattingen die er bestaan over stijl en in 2.3 wordt de opvatting besproken die ten grondslag ligt aan de probleemstelling in deze studie, namelijk dat stijl een afwijking van een bepaalde norm is. Tegen deze stijlopvatting zijn nogal wat bezwaren aangevoerd; deze komen in 2.3 ter sprake.

In 2.4 wordt gesproken over de methode van onderzoek, de manier waarop de gegevens zijn verzameld, de zogenaamde kwantitatieve methode. In deze paragraaf komen de bezwaren tegen deze methode aan de orde en wordt nader ingegaan op het beperkte bereik ervan. De kwantitatieve methode wordt verder uitgewerkt in hoofdstuk 4, Woordfrequenties.

In dit hoofdstuk over stilistiek komen dus alleen stijlopvattingen aan de orde en een methode om stilistische verschijnselen te beschrijven. Maar een andere, heel belangrijke, kwestie is het verklaren van de even-tueel te vinden verschillen. In deze studie wordt daar geen aandacht aan besteed. Hier wil ik er in het kort, met veel literatuurverwijzingen, het volgende over opmerken.

Het verklaren van gevonden verschillen kan niet los gezien worden van het doel of het object van de stilistiek. Redelijk gangbaar als onderzoeksobject in de stilistiek is: het verklaren van talige kenmerken door buitentalige factoren zoals 'onderwerp van de tekst', 'relatie tussen zender en ontvanger' enz. Maar eigenlijk komen we hiermee

weinig verder. We zitten nog steeds op het punt zoals dat is omschreven door *Crystal* (1972). Voordat ik hierop inga, geef ik eerst een lang citaat van dezelfde *Crystal* die drie jaar daarvoor, samen met *Davy*, 'Investigating English Style' publiceerde.

"... the aim of stylistics is to analyse language habits with the main purpose of identifying, from the general mass of linguistic features common to English as used on every conceivable occasion, those features which are restricted to certain kinds of social context; to explain, where possible, why such features have been used, as opposed to other alternatives; and to classify these features into categories based upon a view of their function in the social context." (1969:10)

Crystal en *Davy* stellen dan op basis van de hypothese dat "any use of language displays certain linguistic features which allow it to be identified with one or more extra-linguistic contexts", het volgende:

"Our task is accordingly, threefold:

We must identify the entire range of linguistic features which people intuitively feel to be stylistically significant, and specify a precise way of talking about them (a metalanguage).

We must outline a method of analysis which will allow us to organise these features in such a way as to facilitate comparison of any one use of language with any other.

We must decide on the function of these features, by classifying them into categories based on the kind of extra-linguistic purpose they have." (aangehaald werk - voortaan a.w. - 11,12)

Maar ... drie jaar later spreekt *Crystal* (1972:112) niet meer over een doel. Hij zegt dan:

"... the main function of stylistics is to facilitate discussion of a text (...), what stylistics can hope to do is provide a systematic account of a text, based on general principles which will facilitate comparison, and a precise terminology, so that his and others' observations can be discussed in an atmosphere of mutual intelligibility."

Deze omslag van een zo ambitieus object naar een "systematic account of a text" is kenmerkend voor de stand van zaken binnen de stilistiek.

Opvallend is wel dat in veel stilistische studies een overzicht wordt gegeven van factoren die invloed kunnen uitoefenen op taalgebruik. Ik noem hier: *Enkvist* (1964:30,31) waar hij spreekt over de "extratextual context"; *Spencer en Gregory* (1964:83-91) met hun "placing of the text"; *Crystal en Davy* (1969:66-79) met hun "dimensions of situational constraint" en *Sandell* (1977:47-69) met zijn "antecedents of style". Vanuit de sociolinguïstiek is hier van belang het SPEAKING-model van

Hymes, behandeld door o.a. *Appel e.a.* (1979:43-47); zie ook bijvoorbeeld *Hartveldt* (1978:94-108).

Maar men komt niet verder dan een vaak verdienstelijk overzicht van grotendeels dezelfde factoren, terwijl nog onduidelijk blijft hoe door middel van die factoren verschillen in taalgebruik kunnen worden verklaard. Ook het zogenaamde register-onderzoek, onderzoek dat zich bezighoudt met de correlatie tussen taalgebruik en situatie, is nog niet veel verder dan hetgeen al vermeld is in het overzichtsartikel van *Haft-Van Rees* (1974); zie ook nog *Kusters* (1977) en *Hartemink en Rijlaarsdam* (1978).

Gezien de talrijke problemen bij het verklaren van verschillen in taalgebruik is in deze studie gekozen voor een opzet waarbij het beschrijven van eventueel te vinden verschillen in taalgebruik centraal staat. Dit beschrijven gebeurt uiteraard wel vanuit een bepaalde opvatting over stijl. Voor een goed begrip van die stijlopvatting is het nodig eerst, in vogelvlucht, een overzicht te geven van een aantal belangrijke stijlopvattingen.

2.2 Stijlopvattingen

De stelling dat er een grote verscheidenheid bestaat aan definities van het begrip stijl, is bijna een stijlfiguur geworden in stilistische studies. Een aardig voorbeeld hiervan vinden we bij *Guiraud*. *Guiraud*' (1954a:6) formuleert het in de inleiding van zijn voortreffelijk overzichtswerk aldus:

"Mais le contenu du mot style est si vaste que, soumis à l'analyse, il éclate en une poussière de concepts autonomes qui, en se réclamant ou sans se réclamer du titre de stylistique ont, avec des bases communes, des domaines et des méthodes tranchés." En in zijn conclusie zegt hij (a.w.: 109): "La tâche la plus urgente de la stylistique est de définir son object, sa nature, ses buts et ses méthodes, à commencer par la notion même du style."

Ruim tien jaar later zegt *Milic* (1967:44) in zijn stilistische studie over Swift berustend: "... style has no clear-cut nucleus but rather an outline of amorphous dimensions which tends to collect meanings within itself. In mentalistic phrase, it is a word trying to express the inexpressible."

En weer tien jaar later probeert *Sanders* (1977:6) allerlei verschillende stijlopvattingen onder een noemer te brengen door de kern van het begrip stijl als volgt te omschrijven: "Dass Stil ein Strukturgefüge

darstellt, ein Netz signifikanter Sprachelemente und ihrer Beziehungen im Text, darf wohl als unbestritten gelten ..." Maar bij zijn definitie van stijl als 'Strukturgefüge' blijkt hoever je daarmee komt (a.w.:86): "... ein Strukturgefüge (...) in dem Art, Funktion, Frequenz, Kombination und Distribution von Stilmitteln und stilneutralen Elementen eine charakteristische Verbindung eingehen." Sanders noemt deze omschrijving zelf een 'Pauschalformulierung'. In zijn bespreking van deze formulering vinden wij dan ook geen enkel duidelijk criterium voor het onderscheid 'Stilmittel' en 'stilneutrale Elemente', laat staan een voorbeeld van een karakteristieke verbinding tussen deze twee.

Het blijkt bijzonder moeilijk het begrip stijl precies te omschrijven. Sinds Guiraud zijn we nog niet veel gevorderd. Maar we krijgen wel enige duidelijkheid wanneer we kijken naar de verschillende opvattingen over het begrip stijl.

Enkvist (1964:10-12) heeft in zijn 'On defining style' een poging gedaan de talrijke verschillende opvattingen te classificeren vanuit acht invalshoeken (hier puntsgewijs weergegeven in zijn formulering):

1. the communication process: a. point of view of the writer, b. characteristics of the text, c. impressions of the reader; 2. objectively verifiable versus subjectively impressionistic statements on style; 3. style as a shell surrounding a pre-existing core of thought or expression; 4. as the choice between alternative expressions; 5. as a set of individual characteristics; 6. as deviations from a norm; 7. as a set of collective characteristics; and 8. as those relations among linguistic entities that are statable in terms of wider spans of text than the sentence.

Als we naast dit overzicht van Enkvist de - daarmee gedeeltelijk overeenkomende - indeling zetten die *Spillner* (1974) tien jaar later geeft voor zijn vierde hoofdstuk in 'Linguistik und Literaturwissenschaft', dan krijgen we een vrij volledig beeld van de verschillende benaderingen binnen de stilistiek:

1. Stil als textimmanentes Phänomen; 2. Stil als rhetorische Ausarbeitung und ästhetische Beigabe; 3. Stil: Reflex der dichterischen Persönlichkeit; 4. Stil als Affektivität und Konnotation; 5. Abweichungsstilistik; 6. Stilauffassungen in der generativen Transformationsgrammatik; 7. Stil als Auswahl; 8. Stil als Kontrast im Kontext; 9. Konstituierung von Stil durch den Leser; 10. Funktionalstilistik; 11. Theorie der sprachlichen Register.

Heeft het, bij zo'n grote verscheidenheid aan opvattingen, nog zin om het begrip stijl te blijven gebruiken? Gray (1969:104) meent van niet. Hij ziet in zijn 'Style, the problem and its solution' als enige oplossing het verwerpen van het begrip stijl:

"Might there not be 'something' to a concept which so many seem agreed upon and which has, indeed, been agreed upon for centuries? The answer to this question is threefold. One, agreement is more likely to be found in theories or programs than in practices. Two, style would not be the first prevalent and enduring concept which has been found to be fallacious. Three, it is fallacious."

Gray vergelijkt het concept stijl met het concept ether. Dit laatste concept was (a.w.:108) "at best a stop-gap hypothesis in lieu of adequate explanations and at worst an inadequate but comfortable explanation which blinded scientists to the need for a re-evaluation of their conclusions." Stijl, aldus Gray (a.w.:109), is "an entity with no attributes of its own, whose existence has never been proven, which is assigned the properties of whatever system it happens to be applied to, and therefore has numerous contradictory properties ..."

Terecht stelt *Enkvist* (1973:13) dat Gray geen onderscheid maakt tussen 'substantive terms' en 'notational terms': "Substantive terms are irreplaceable without loss of conceptual content, whereas notational terms are, basically, abbreviations and can thus be paraphrased."

Maar hoe nu verder? Wat te doen met al die stijloppvattingen, al die parafrases van een notationale term? Het lijkt mij niet zinnig om de voor- en nadelen van alle parafrases tegen elkaar af te wegen. Dit nog afgezien van het feit dat in het bestek van deze studie aan geen ervan voldoende recht zou kunnen worden gedaan. Zinniger lijkt mij een op de praktijk gerichte studie met de volgende werkwijze: uitgaande van een bepaalde stijloppvatting, namelijk stijl als afwijking van een (verwante) norm, probeer ik de bezwaren tegen deze opvatting te weerleggen (hierover meer in 2.3), en ik hoop dan door onderzoek naar verschillen in taalgebruik aan te tonen dat die stijloppvatting vruchten afwerpt.

Tot nu toe heeft namelijk *Gray*, de auteur die beweert dat stijl niet bestaat, gelijk met zijn voor mij stimulerende opmerking (a.w.:98):

"One important criticism that can be made of the stylistically-minded linguists on both sides of the Atlantic is that they have never clearly demonstrated that style exists and is something that can be studied. They have not studied it; they have only talked about studying it."

Gray had zich wel iets voorzichtiger mogen uiten na de bundel 'Style in Language' (1960) en na bijvoorbeeld *Milic's* 'A Quantitative Approach

to the Style of Jonathan Swift' (1967). Maar ik ben het wel met hem eens dat de theoretische bespiegelingen in stilistische literatuur in geen verhouding staan tot het onderzoek. Overigens komt deze kritiek niet alleen van iemand die het begrip stijl wil laten vallen. Ook *Crystal* (1972:104) signaleert dat "...theorizing has gone far ahead of experimental evidence, and as a result, pseudo-procedures and pseudo-problems have multiplied." "... a concern to think in terms of rigorous hypothesis-testing (...) is sadly lacking in the published literature."

2.3 Stijl als afwijking van een norm

Wanneer we stijl opvatten als afwijking van een norm, rijzen er enkele problemen, die vooral te maken hebben met de moeilijkheden bij het definiëren van de norm. Ook het definiëren van de afwijking geeft problemen, maar deze kunnen beter behandeld worden wanneer er meer duidelijkheid is over de norm.

2.3.1 De norm?

Wanneer iemand een bepaalde manier van taalgebruik 'moeilijk' of 'vreemd' of 'ouderwets' vindt, dan vindt hij dat kennelijk omdat die vorm van taalgebruik anders is dan (afwijkt van) de een of andere norm, uiteraard impliciet, die aangeeft welk taalgebruik 'gemakkelijk', 'gewoon' of 'modern' is. Zo'n norm is dan de, hypothetische, in de taalgebruiker aanwezige maatstaf waaraan hij afmeet in hoeverre een gegeven vorm van taalgebruik 'normaal' of 'abnormaal' is. Het lijkt overigens zeer aannemelijk dat deze impliciete norm verschilt per taalgebruiker. En het zal duidelijk zijn dat deze norm zich vooralsnog onttrekt aan wetenschappelijk onderzoek.

Wanneer echter stilistisch onderzoek wordt gedaan op basis van stijl, gedefinieerd als afwijking van een norm, is het bijzonder belangrijk dat die norm zo nauwkeurig mogelijk wordt aangegeven. Hoe de norm gekozen wordt, is wetenschappelijk gezien van minder belang. De onderzoeker kan bijvoorbeeld trachten zijn eigen impliciete norm te expliciteren, maar ook een speciaal voor dit doel geconstrueerde norm kan bruikbaar zijn. Uiteraard is het gewenst dat zo'n geconstrueerde norm enigszins aansluit bij wat taalgebruikers als 'normaal' taalgebruik ervaren, maar, nogmaals, wetenschappelijk is vooral een nauwkeurige definitie van belang. Pas dan kan onderzoek immers leiden tot wetenschappelijk

verantwoorde, verifieerbare, resultaten. Bij stilistisch onderzoek zal deze werkwijze leiden tot duidelijke conclusies over stilistische (want als zodanig gedefinieerde) afwijkingen in een gegeven groep teksten.

Maar nu, welke norm moet de onderzoeker kiezen? Hieronder volgen een aantal suggesties die in de literatuur over stilistiek veel aandacht hebben gekregen. Helaas is het bij suggesties gebleven. En dat is ook niet verwonderlijk, want elk voorstel voor een norm roept grote problemen op.

Het zal duidelijk zijn dat een norm als het 'normale' taalgebruik of het 'juiste' taalgebruik grote definitie-problemen geeft. Het zal ook duidelijk zijn dat het ondoenlijk is om als norm het taalsysteem - wat men daar ook onder wil verstaan - of alle taaluitingen te nemen. Dit laatste doet de veel geciteerde *Bloch* (zie bijvoorbeeld *Freeman* 1970:5 en *Saporta* 1975:87): "The style of a discourse is the message carried by the frequency-distributions and transitional probabilities of its linguistic features, especially as they differ from those of the same features in the language as a whole." Maar de bedoelde gegevens van 'the language as a whole' zijn nog bij lange na niet bekend. Terzijde zij nog vermeld dat *Riffaterre* (1960) heeft geprobeerd de problemen te omzeilen door de context als norm te nemen. Maar zijn voorstel heeft heel terecht weinig positieve reacties gekregen (zie o.a. *Carstensen* 1970:259), hoewel *Sanders* (1977:67-71) nog op *Riffaterre*'s voorstel voortborduurde door in zijn stilistische benadering vanuit de lezer, de termen 'Erwartungsbruch' en 'Erwartungserfüllung' te hanteren.

Een andere interpretatie van stijl als afwijking formuleert *Sandell* (1977:12). Hij stelt voor "to describe a style, in its different aspects, in terms of how much it deviates from the average in the sample of which it characterizes a unit, using, for example, the standard deviation as unit. Here, 'norm' means 'average'." Maar deze werkwijze kan alleen gevolgd worden, wanneer men bij een steekproef van een bepaald type taalgebruik wil weten in hoeverre een bepaalde tekst uit die steekproef afwijkt van het gemiddelde van die steekproef. Deze werkwijze kan dus niet toegepast worden wanneer men wil weten waarin een bepaalde groep teksten verschilt van een andere groep.

Enkvist (1964:28) heeft een nadere interpretatie van het begrip norm gegeven door in zijn definitie het begrip 'contextually related norm' op te nemen: "The style of a text is a function of the aggregate of the ratios between the frequencies of its phonological, grammatical

and lexical items, and the frequencies of the corresponding items in a contextually related norm." Over het begrip 'context' blijft Enkvist bijzonder vaag (a.w.:29): "... another text or corpus which is regarded as a norm and which has a definite contextual relationship ...". Merk op dat hier het begrip 'context' gedefinieerd (?) wordt met behulp van het woord 'contextual'. Een fraai voorbeeld van een cirkeldefinitie. Ook over die 'contextual relationship' blijft Enkvist vaag. Blijkens zijn schema (a.w.:30,31) bevat die relatie alle tekstuele verschijnselen (vanaf fonetische tot compositorische) en alle extratekstuele verschijnselen (periode, genre, auteur, lezer, situatie, gebaren enz.). In een latere publikatie van Enkvist (1973:145) vinden we dezelfde cirkeldefinitie terug: "context being definable as the aggregate of stylistically relevant features in the textual and situational envelopes surrounding the text." (onderstreept van mij) Zo wordt stijl gerelateerd aan een norm, die zelf weer gedefinieerd wordt met behulp van het woord 'stilistisch'.

Desondanks lijkt het idee van de verwante norm het meest vruchtbaar. Het is pas zinvol over kenmerken van een groep teksten te spreken, wanneer een verwante groep teksten die kenmerken niet of in mindere mate heeft. Het probleem blijft echter: wat is een verwante groep teksten?

In deze studie is voor de taal van 'Den Haag' als verwante norm gekozen het taalgebruik in kranten en populair-wetenschappelijke publikaties. Deze keuze kan verdedigd worden met citaten uit het volgende hoofdstuk, Oordelen over de taal van 'Den Haag', waarin vergelijkingen worden getrokken tussen het taalgebruik in overheidsteksten enerzijds, en kranten en wetenschappelijke publikaties anderzijds. Verder kan gewezen worden op het feit dat al deze typen teksten gekarakteriseerd kunnen worden als informatieve teksten. Maar belangrijker dan deze rechtvaardiging op basis van intuïtie, is het argument dat het taalgebruik in kranten en populair-wetenschappelijke publikaties al is onderzocht volgens dezelfde methode die in deze studie gehanteerd wordt (zie pagina 4). Ik doel hier op de corpora 'dagbladen', 'opiniebladen' en 'populair-wetenschappelijke boeken' uit *Woordfrequenties*. De vergelijking werd zo aanzienlijk vergemakkelijkt. Voor deze studie behoefde alleen een corpus 'overheidsteksten' te worden geanalyseerd.

2.3.2 Bezwaren tegen de opvatting van stijl als afwijking van een norm

In de literatuur over stilistiek zijn talrijke bezwaren opgesomd tegen deze stijlopvatting. Zie bijvoorbeeld Enkvist (1973:98-108), *Spillner*

(1974:39-40) en Sandell (1977:10-12). Ik beperk me hier tot Spillner omdat hij mijns inziens de bezwaren het meest overzichtelijk opsomt. Spillner noemt zeven bezwaren en concludeert dan zonder verder commentaar: "Die Konzeption von Stil als Abweichung ist als heuristisches Mittel geeignet, nicht dagegen als Grundlage einer Stiltheorie." Hier volgen de bezwaren, voorzien van mijn commentaar.

1. "Aus der Theorie würde sich ergeben, dass es Texte ohne Stil gibt (d.h. diejenigen Texte, die nicht von einer Norm abweichen)."

Als iemand uitgaat van de opvatting dat stijl bestaat uit een afwijking van een norm, en bepaalde teksten of tekstgedeelten als norm kiest, dan lijkt het inderdaad of hij uitgaat van de vooronderstelling dat die norm geen stijl heeft of tenminste stilistisch als neutraal te beschouwen is. Hetzelfde geldt voor teksten die geheel gelijk zijn aan de norm. Dit is echter alleen maar een kwestie van definiëren. De onderzoeker kiest een vergelijkingspunt, in dit geval een groep teksten die hij als norm neemt en die hij daarom - gemakshalve - stilistisch neutraal noemt. Maar dit wil beslist niet zeggen dat die norm-teksten geen stijl hebben.

2. "Weder Norm noch Abweichung sind exakt definierbar."

In de vorige paragraaf is al aangegeven, dat het wèl mogelijk is een norm te definiëren. Deze norm is een wetenschappelijke constructie, maar verder kunnen wij nu eenmaal niet gaan, willen wij wetenschappelijk verantwoorde uitspraken kunnen doen. Bij elke stilistische probleemstelling kan een norm vastgesteld worden en - bijgevolg - een maat voor de afwijking van die norm. Een voorbeeld: Voor een onderzoek naar stijlverschijnselen in de taal van 'Den Haag' kan men als norm kiezen het taalgebruik in kranten en populair-wetenschappelijke publikaties. Men zou dan de stilistische afwijking in de taal van 'Den Haag' kunnen definiëren door, bijvoorbeeld bij verschijnselen die te maken hebben met het voorkomen van bepaalde woordsoorten, gebruik te maken van de toetsingstheorie uit de statistiek (zie pagina 133 - 136). Uiteraard blijft het de vraag of de gekozen norm en afwijking als zodanig aanvaardbaar zijn, maar dit is een heel andere kwestie. Zowel norm als afwijking zijn nauwkeurig omschreven door middel van een operationele definitie.

3. "Stil wird rein negativ definiert, ohne dass sich daraus in der Stiltheorie qualitative Aussagen ergeben."

Dit is een heel merkwaardige opvatting! Waarom moet een afwijking als negatief worden beschouwd?

4. "Die Menge der Abweichungen und die Menge der Stilmerkmale sind weder deckungsgleich, noch schliesst eine der beiden Mengen die andere ein. Es gibt Abweichungen ohne Stileffekt (...). Andererseits ist anzunehmen, dass auch nichtabweichende sprachliche Elemente stilistisch relevant sein können."

Dit bezwaar kan pas in de discussie worden betrokken wanneer duidelijk is wat er onder 'Stileffekt' of 'stilistisch' moet worden verstaan. Juist bij het bekritisieren van een stijloppvatting moet duidelijk zijn wat er bedoeld wordt met stijl. Overigens houdt de opvatting van stijl als afwijking van een norm volstrekt niet in, dat daarmee alle verschijnselen die men ooit onder 'stijl' heeft willen vatten, beschreven kunnen worden.

5. "In kommunikationstheoretischer Sicht ist zu kritisieren, dass die Kategorien 'Autor' und 'Leser' zugunsten des Vergleichs eines Textphänomens mit der Norm vernachlässigt sind."

Onnodig te zeggen dat een onderzoeker die uitgaat van stijl als afwijking van een norm, daarmee niet direct het doel heeft een bijdrage te leveren aan de communicatietheorie, zoals bedoeld door Spillner. Maar dit wil niet zeggen dat er op basis van zo'n studie niet onderzocht kan worden in hoeverre een auteur afwijkt van een norm of in hoeverre lezers een tekst beoordelen als afwijkend van een norm.

6. "Die Theorie ist literaturwissenschaftlich anwendbar nur für einen sehr idiosynkratischen Stil und besondere poetische Experimente mit der Sprache, nicht dagegen bei Autoren, die einen 'normalen' Stil schreiben." Deze opmerking is onjuist, en niet alleen omdat Spillner het hier heeft over literaire teksten. Ook wanneer we stijlen willen onderzoeken die minder idiosyncratisch, dus meer algemeen zijn, is het best mogelijk een norm te kiezen en een maat voor afwijking daarvan te bepalen, opdat de desbetreffende stijl in kaart gebracht kan worden. Zie verder het commentaar bij het tweede bezwaar.

7. "Es besteht die Gefahr, dass bei einer literaturwissenschaftlichen Interpretation nur die wenigen ungewöhnlichen Stilmerkmale berücksichtigt werden bei Vernachlässigung des ganzen Textes und seiner Struktur." Zelfs al laten we deze tegenwerping, die blijkbaar gericht is op de literatuurwetenschap, slaan op onderzoek naar niet-literaire teksten, dan nog is zij gemakkelijk te ontzenuwen. Dat er voor de structuur van de tekst weinig aandacht bestaat, is te verklaren uit het feit dat het onderzoek op dit punt nog in de kinderschoenen staat.

Wanneer de structuur van een tekst voorwerp wordt van stilistisch onderzoek, is het, bij 'stilistisch' opgevat als 'afwijkend', heel wel mogelijk een bepaalde structuur als norm te nemen en na te gaan in hoeverre de te onderzoeken teksten hiervan afwijken. Tot zover het commentaar op de bezwaren tegen stijl als afwijking van een norm.

2.4 De kwantitatieve methode

In de literatuur worden twee vormen van stilistisch onderzoek nog vaak ten onrechte tegenover elkaar gesteld. Enerzijds het linguïstisch georiënteerde stilistische onderzoek, dat tot doel heeft objectief-verifieerbare gegevens te verkrijgen over eigenschappen van teksten met behulp van frequentietabellen en significantietoetsen; de kwantitatieve benadering. Anderzijds de kwalitatieve benadering: het uit een rijke literaire traditie stammende subjectief-impressionistische onderzoek. Veel beoefenaren van de kwalitatieve stilistiek staan sceptisch tegenover wat zij beschouwen als een reducerende en meestal overbodige inbreng vanuit de linguïstiek.

Aangezien in deze studie wordt beschreven hoe een stilistisch probleem wordt benaderd via de kwantitatieve methode, zullen een aantal bezwaren tegen de kwantitatieve stilistiek hier nader worden onderzocht. Eerst een aantal extreme opvattingen over de kwantitatieve methode; polariserende opvattingen, die voor- en tegenstanders van de kwantitatieve benadering verder uit elkaar drijven dan nodig is, en die bovendien laten zien dat er nog veel onduidelijkheid bestaat over de aard en het bereik van deze methode (2.4.1). Daarna komen de bezwaren aan de orde die Ullmann (1964) ertegen heeft aangevoerd (2.4.2). Ik beperk me tot Ullmann omdat hij het meest uitgebreid de bezwaren opsomt en omdat zijn argumenten nu nog steeds worden aangehaald. In paragraaf 2.4.3 wordt duidelijk gemaakt dat de kwantitatieve en de kwalitatieve methode minder van elkaar verschillen dan wel wordt gesuggereerd. In die paragraaf zal tot slot nog iets worden gezegd over het presenteren van kwantitatieve gegevens.

2.4.1 Polariserende opvattingen

Uit de volgende vijf opvattingen wordt duidelijk hoe verschillend er gedacht wordt over de kwantitatieve methode. Eerst komen een te positieve en een te negatieve opvatting ter sprake; daarna drie opvattingen die beter gekwalificeerd kunnen worden als misvattingen.

1. *Guiraud* (1954a:106) stelt dat de stilistiek zich bij uitstek leent voor een statistische benadering: "La stylistique semble en fait le domaine d'élection de l'analyse statistique; non seulement parce que les faits y sont objectivement observables et dénombrables, mais parce que la langue est une entité statistique, 'une somme d'empreintes', et c'est l'emploi plus ou moins généralisé d'une expression par telle ou telle catégorie d'usagers qui crée sa valeur stylistique."

Hier ziet *Guiraud* mijns inziens over het hoofd, dat lang niet alle stilistische elementen zich goed lenen voor kwantificering en dus voor een statistische benadering. Alleen stilistische elementen die nauwkeurig omschreven kunnen worden, zijn via de kwantitatieve methode te onderzoeken, bijvoorbeeld verschijnselen als woord- en zinslengte. Maar diezelfde methode is minder bruikbaar bij een stilistisch onderzoek naar verschijnselen die moeilijker zijn te omschrijven, bijvoorbeeld beeldspraak of compositie.

2. Een voorbeeld van een te negatieve benadering van de kwantitatieve stilistiek is te vinden in *Gray* (1969:21): "Yet it is not clear that statistics has, thus far at least, contributed anything positive to establishing the existence of such an entity as style." *Elders* (a.w.: 42) spreekt *Gray* zelfs over: "the cold dead hand of statistics." Het is jammer dat *Gray* de publikatie van *Milic* (1967) niet in zijn beschouwingen heeft betrokken. *Milic* heeft mijns inziens zeker het doel bereikt dat hij voor ogen had met zijn kwantitatief-stilistische studie over *Swift*, namelijk (a.w.:270): "... to steer the study of style away from vagueness and impressionism and towards the maximum of objectivity and precision consistent with the nature of the problem."

3. *Spencer en Gregory* (1964:84) beweren het volgende naar aanleiding van een linguïstische beschrijving met statistische gegevens: "... the function of linguistic description is not simply that of making the precise nature of (...) features explicit, or of producing statistical tables to support the intuitive judgement. Such procedure involves the danger of fixing the significant stylistic features in advance of careful analysis, thereby closing the door against the possibility of modifying and developing the original 'hypothesis'."

Uiteraard kan cijfermateriaal misbruikt worden, in die zin dat een onderzoeker al te gemakkelijk vaststelt dat zijn hypothesen worden bevestigd door kwantitatief onderzoek, zonder zich af te vragen of de hypothesen niet onvolledig of triviaal waren, en daarmee de deur sluitend

voor verder onderzoek. Maar dit bezwaar geldt elk onderzoek, ook het kwalitatief-stilistische onderzoek. Het bezwaar is alleen te ondervangen door inzicht te geven in de hypothese-vorming en door een discussie naar aanleiding van de onderzoeksresultaten.

Gelukkig geven Spencer en Gregory op een andere plaats (a.w.:91) wel een juiste visie op de kwantitatieve benadering vanuit de linguïstiek:

"A detailed analysis of linguistic features within the text has as one of its aims to cut beneath the generalizations, to get behind the metaphorical labels, of which the literary study of style makes such use."

4. Een ander voorbeeld van een onjuiste benadering geeft *Kroeber* (1969: 207) wanneer hij spreekt over de reductie die plaatsvindt wanneer we een literaire tekst opvatten als een tekst die bestaat uit taalelementen: "Quantitative analysis tends to make this fatal reduction; by discriminating parts it blurs the qualitative relationship of part to whole which in fact defines the essential nature of the part."

Kroeber suggereert hier eigenlijk - hij geeft geen nadere uitleg - dat er bij een kwalitatieve analyse, in tegenstelling tot een kwantitatieve analyse, geen gevaar voor reductie bestaat. Maar ook in de kwalitatieve analyse vindt reductie plaats doordat een aantal, om wat voor reden dan ook opvallende, elementen ter nadere bestudering uit het geheel van een tekst gelicht worden. Men kan hier nog tegenwerpen dat een kwalitatieve analyse minder of op een andere manier reduceert dan een kwantitatieve analyse. Maar dan nog zou de laatste vorm van onderzoek de voorkeur verdienen, omdat daarbij objectief-verifieerbare gegevens gepresenteerd worden, en dat is bij de huidige stand van stilistisch onderzoek hoogst noodzakelijk.

5. De derde misvatting komen we tegen bij *Sandell* (1977:23). Ik ga hier voorbij aan de opmerking van Sandell dat de toepasbaarheid en relevantie van kwantitatief onderzoek beperkt zijn, omdat het tellen altijd "a preoccupation with surface phenomena in texts" tot gevolg heeft. Want een kwalitatieve benadering houdt zich ook bezig met 'surface phenomena'. Dit blijkt uit wat Sandell verder nog op dezelfde pagina zegt. Na een vage aanduiding van de transformationeel-generatieve benadering in de stilistiek schrijft hij, dat stijl waarschijnlijk inderdaad "a matter of surface appearance" is. Met dit in- en uitpraten komen we weinig verder.

Belangrijker is de opmerking van Sandell dat veel kwantitatieve studies niet verder komen dan een lijst met parameters van een bepaalde populatie zonder hieraan een interpretatie toe te voegen. Maar ook het geven

van parameters zonder interpretatie is een legitiem, zij het een beperkt, doel. Een voorbeeld: teksten kunnen op basis van deze parameters veranderd worden om te onderzoeken of lezers dan anders oordelen over die teksten. En Sandell omzeilt het probleem dat bij een kwalitatieve benadering de interpretatie of de waardering van stilistische elementen subjectief blijft, als er geen onderzoek wordt gedaan naar de betrouwbaarheid van die interpretatie of waardering.

Na deze bespreking van vijf, mijns inziens, onjuiste opvattingen over de aard en het bereik van de kwantitatieve methode, volgen hier nog de kernpunten van mijn commentaar voordat de meer wezenlijke bezwaren van Ullmann aan de orde komen. Een kwantitatieve analyse heeft alleen maar zin als de te onderzoeken verschijnselen nauwkeurig kunnen worden omschreven. Omdat dit bij de huidige stand van onderzoek lang niet altijd het geval is, blijft de waarde van een kwantitatieve analyse beperkt. Daaraan voorafgaand en daarnaast behoudt een kwalitatieve analyse haar waarde. Maar het voordeel van een kwantitatieve analyse is, dat er objectief-verifieerbare gegevens te voorschijn komen, die de vaagheden van een kwalitatieve analyse kunnen verhelderen. Zo kan men bijvoorbeeld een poging wagen subjectief-impressionistische oordelen over teksten te relateren aan linguïstische eigenschappen van die teksten.

2.4.2 Bezwaren tegen de kwantitatieve methode

De belangrijkste bezwaren tegen een kwantitatieve benadering in de stilistiek zijn naar voren gebracht door Ullmann (1964:118,119). Hij noemt weliswaar alleen de statistische methode, maar uit zijn opmerkingen blijkt dat hij ook kwantificering zonder statistische bewerking op het oog heeft. Zijn bezwaren, voorzien van mijn commentaar, zijn de volgende:

1. "The statistical method is too crude to catch some of the subtle nuances of style (...)."

Inderdaad. Maar dit neemt niet weg dat een kwantitatieve analyse wel geschikt kan zijn om hypotheses over een reeks betrekkelijk eenvoudige stilistische verschijnselen te toetsen. Een volledige analyse die ook subtiele stijlnuanceringen omvat, is hierbij niet het doel. Vergelijk in dit verband weer Milic (1967:270) die aan het einde van zijn kwantitatief stilistische studie over Swift het volgende opmerkt: "I have not claimed ... to provide a complete description of Swift's style and a foolproof method of attribution. I do not know that either is possible." Dat een complete beschrijving onmogelijk is, zal zonder meer duidelijk zijn.

2. "Numerical counts may impose a spurious kind of precision on data too intricate or too fluid to admit of such treatment."

Inderdaad! Maar ligt dit aan de methode of aan de manier waarop de methode wordt toegepast en gepresenteerd? Wanneer een onderzoeker zonder nader commentaar frequentietabellen geeft van verschijnselen die te gecompliceerd zijn voor de gehanteerde meetmethode, ligt de fout bij de onderzoeker en niet bij de methode. Uit het onderzoeksverslag moet duidelijk blijken dat de te onderzoeken verschijnselen ook werkelijk gemeten zijn, en - wat even belangrijk is - juist gemeten zijn. Vooral hier wordt het beperkte bereik van de kwantitatieve stilistiek duidelijk. Maar 'beperkt bereik' is niet hetzelfde als 'geen bereik'.

3. "One great weakness of the so-called 'stylostatistical' method is that it makes no provision for the influence of context (...)."

Deze opmerking is onjuist. Milic (1967) onderzocht al drie-woordpatronen. En bij het onderzoek waarvan in deze studie verslag wordt gedaan, kan van elk woord de context worden opgezocht door middel van een concordantielijst. Als men bijvoorbeeld wil nagaan of het woord 'aanzien' alleen maar voorkomt in de voorzetseluitdrukking 'ten aanzien van', dan kost dat weinig moeite. En bovendien is het lang niet altijd nodig de context in het onderzoek te betrekken. Wanneer men wil weten hoe vaak zelfstandige naamwoorden eindigend op '-heid' voorkomen, of wanneer men als stilistische parameter de verhouding tussen het aantal bijvoeglijke en zelfstandige naamwoorden kiest, dan heeft men voldoende aan frequentielijsten zonder informatie over de context.

4. "Another danger inherent in this approach is that quality is overshadowed by quantity and very diverse elements are lumped together on the ground of some superficial similarity."

Dit gevaar is inderdaad erg groot, maar het kan voorkomen worden. Aan het eind van de vorige paragraaf is al gesteld dat een kwalitatieve analyse haar waarde houdt naast een kwantitatieve benadering. Deze laatste benadering kan hoogstens getalsmatige gegevens opleveren over linguïstische elementen die om wat voor reden dan ook stilistisch interessant lijken. Waarschijnlijk - Ullmann geeft geen nadere uitleg - wordt met het tweede gedeelte van het bezwaar onder andere bedoeld op homonymie-verschijnselen. Bijvoorbeeld, dat bij een onderzoek naar het voorkomen van de betrekkelijke voornaamwoorden 'die' en 'dat', ook de aanwijzende voornaamwoorden 'die' en 'dat' worden meegeteld. Het is uiteraard een groot gebrek van de kwantitatieve benadering, als met dit type verschijnselen geen rekening wordt gehouden. Maar, nogmaals, dan is dat een fout van de onderzoeker, niet een gebrek van de methode. Bij

de bespreking van het in deze studie toegepaste codeersysteem komen homonymie-verschijnselen nog uitvoerig aan de orde.

5. "It also happens that an imposing array of figures produces a result which one could have noticed with the naked eye or which is so obvious that it needs no demonstration."

Inderdaad, als men de studies van *Herdan* (1956, 1962, 1966) heeft door- geworsteld, moet men concluderen dat Herdans statistische benadering indrukwekkend is, maar gezien de huidige stand van onderzoek nog weinig relevantie heeft voor de kwantitatieve stilistiek. Maar laten we ons nu beperken tot het onderzoek dat in deze studie is uitgevoerd, name- lijk het opsporen van linguïstische kenmerken in een bepaalde categorie teksten (de correspondentie tussen regering en Staten-Generaal), naar aanleiding van subjectief-impressionistische oordelen over de taal van 'Den Haag'. Als het lukt om via de kwantitatieve benadering een aantal linguïstische kenmerken op het spoor te komen, kan een voorstander van de kwalitatieve benadering zeggen dat het resultaat triviaal is of al lang duidelijk was. Maar hierbij wordt dan over het hoofd gezien dat het doel van het onderzoek juist is: het objectiveren en verifiëren van subjectief-impressionistische oordelen. En dan zal uit deze studie blijken dat sommige intuïtief 'vastgestelde' eigenschappen van teksten helemaal niet zo duidelijk zijn dat ze geen bewijs meer nodig hebben. Integendeel, een aantal intuïtieve oordelen over linguïstische ken- merken blijkt niet bevestigd te kunnen worden voor de onderzochte teksten.

Uit bovenstaand commentaar op de bezwaren van Ullmann moge duidelijk worden, dat niet elke kwantitatieve benadering valt onder de methode die Ullmann bekritiseert. Deze studie heeft niet als doel subtiele stijlnuances op het spoor te komen. Verder kan, indien nodig, de con- text van een woord in het onderzoek worden betrokken. Ook zal bij het bespreken van het codeersysteem blijken, dat er geen verschijnselen onder één noemer worden gebracht die alleen oppervlakkig gezien gelijk zijn. Bovendien zal, bij het opstellen van hypotheses en de discussie naar aanleiding van de resultaten, worden aangegeven hoe nauwkeurig een bepaald verschijnsel kan worden onderzocht binnen het bereik van de in deze studie gehanteerde kwantitatieve methode. En het zal duide- lijk zijn dat het produceren van tabellen alleen als doel heeft over- zichtelijk aan te geven of een intuïtief oordeel gerelateerd kan wor- den aan linguïstische kenmerken. Wel blijft het bezwaar van Ullmann staande dat kwaliteit kan worden overschaduwed door kwantiteit. In de volgende paragraaf wordt hierop nader ingegaan, nadat is aangetoond

dat de kwalitatieve en kwantitatieve methode minder van elkaar verschillen dan wordt gesuggereerd door de auteurs, die zijn aangehaald in deze en de vorige paragraaf.

2.4.3 Algemene opmerkingen over de kwantitatieve methode

In deze laatste paragraaf over de kwantitatieve versus de kwalitatieve methode in de stilistiek, komen vier punten ter sprake, die zelden worden genoemd bij de bespreking van de kwantitatieve benadering. Daarna volgen nog enkele opmerkingen over het presenteren van kwantitatieve en statistische gegevens.

Maar al te vaak zien critici van de kwantitatieve benadering over het hoofd, dat ook bij kwantitatief onderzoek veel 'kwalitatieve' beslissingen genomen moeten worden. De hieronder genoemde beslissingen zullen in de volgende hoofdstukken worden verantwoord.

1. De onderzoeker moet op basis van eigen of andermans intuïtieve oordelen hypothesen opstellen en die zo formuleren, dat ze kwantitatief te verwerken zijn. Met andere woorden: hij bepaalt welke linguïstische elementen onderzocht worden en hij bepaalt de manier waarop.
2. De onderzoeker bepaalt welke teksten onderzocht worden. Hij moet aanvaardbaar maken, dat de gekozen teksten geschikt zijn om de hypothesen te toetsen, die hij heeft opgesteld over het type taalgebruik waartoe hij die teksten rekent.
3. De onderzoeker beslist hoe de steekproef wordt getrokken. Als hij deze beslissing niet verantwoordt, is dat niet alleen een slordige presentatie, maar ook een bewijs van onkunde over de invloed van de steekproeftrekking op het uiteindelijk geselecteerde onderzoeksmateriaal.
4. De onderzoeker moet vaststellen of, en zo ja wanneer, een bepaald linguïstisch element kenmerkend is voor een bepaalde tekst of categorie teksten.

Naar aanleiding van de opmerking van Ullmann over de "imposing array of figures" in publikaties over kwantitatief-stilistisch onderzoek, vermeld in de vorige paragraaf, tot slot nog iets over het presenteren van kwantitatieve en statistische gegevens. De tabellen en uitkomsten van toetsen zijn niet bedoeld om indruk te maken. Zij dienen slechts voor een precieze en volledige weergave van objectief-verifieerbare gegevens.

Daarom is in deze studie niet de manier van presenteren gehanteerd die *Crystal en Davy* (1969) hebben toegepast. Zij hebben ervan afgezien om precieze kwantitatieve en statistische informatie te geven, omdat het toepassen van een statistische toets te gedetailleerd was voor hun doelstellingen. Goed, laat dat zo zijn, maar daarna maken *Crystal en Davy* een vreemde gedachtensprong als zij zeggen (a.w.:22):

"Consequently we make use of the alternative method of expression, making our statements of frequency in more informal terms, using such quantifiers as 'rarely', 'commonly', 'often', 'very often', and so on. This range of adverbials in English cuts up a continuum of frequency very clearly, with very little overlap, and is readily intelligible. It should not be forgotten, however, that any statements of relative frequency in these terms can be referred if necessary to the precise statistical situation which underlies them."

Hier zitten wel addertjes onder het gras! Ten eerste, hebben de genoemde 'quantifiers' wel zo'n duidelijk onderscheiden betekenis? En, als ze toch teruggevoerd kunnen worden op de 'precise statistical situation', waarom dan niet de getallen tussen haakjes vermeld? Ten tweede, deze manier van presenteren kan een vertekend beeld geven. Het blijft zo onduidelijk in hoeverre het toekennen van 'quantifiers' beïnvloed wordt door het verwachtingspatroon dat de onderzoeker van de tekst heeft. Bijvoorbeeld wanneer iemand zou beweren, dat enkele kenmerken van archaïsch taalgebruik 'veel' voorkomen in damesbladen en ambtelijke teksten, dan kan deze bewering gebaseerd zijn op getallen die verschillend zijn voor deze twee categorieën teksten. Immers archaïsch taalgebruik in damesbladen zal waarschijnlijk meer opvallen dan archaïsch taalgebruik in ambtelijke teksten. Daarom is het in alle gevallen beter, in de presentatie de getallen te vermelden waarop oordelen als 'zelden', 'vaak' enz. zijn gebaseerd.

Bij het toetsen van hypothesen over de taal van 'Den Haag' zullen dan ook de frequentietabellen en de uitkomsten van de toetsen worden gepresenteerd. Pas dan is controle op en een zinvolle discussie over de uitkomsten mogelijk.

Maar nu eerst, in het volgende hoofdstuk, een overzicht van de kwalitatieve, ofwel subjectief-impressionistische oordelen over de taal van 'Den Haag', die volgens de in deze studie gehanteerde kwantitatieve methode op hun geldigheid zullen worden getoetst.

3 OORDELEN OVER DE TAAL VAN 'DEN HAAG'

3.1 Inleiding

Dit hoofdstuk bevat een overzicht van oordelen over de taal van 'Den Haag'. Zonder schifting zijn opmerkingen genoteerd over het taalgebruik in teksten van de overheid. Het geven van citaten wil overigens lang niet altijd zeggen dat ik het met de inhoud ervan eens ben. Het gaat in dit hoofdstuk alleen om een overzicht. In deze inleiding wordt eerst iets verteld over de manier waarop de oordelen zijn verzameld en over de beperkingen die ik mij daarbij heb opgelegd (3.1.1). Daarna volgen in 3.1.2 nog enkele opmerkingen over de opbouw van dit hoofdstuk.

3.1.1 Beperkingen

De oordelen zijn afkomstig uit: publikaties over verzorgd Nederlands, taalbeheersing, bureaucratie of vierde macht; lezingen over de taal van de ambtenaar; kranteartikelen en rubrieken in personeelsbladen van departementen en gemeenten over dit onderwerp; taalcursussen voor ambtenaren en officiële aanwijzingen voor het taalgebruik in ambtelijke stukken. Veel materiaal was te vinden in het archief van de op pagina 1 vermelde Commissie Duidelijke Taal.

Vaak werden in plaats van de omschrijving "de taal van 'Den Haag'" andere termen gebruikt: administratieve stijl, ambtelijke stijl, ambtenaren-taal, 'burokraats', CRM-jargon, kanselarijstijl, 'officiaalees', overheidstaal, referendarissenproza, regeertaal en stadhuistaal. De hier gegeven termen worden als synoniemen beschouwd. In deze studie wordt gemakshalve gesproken over de taal van 'Den Haag'.

Bij het verzamelen van oordelen heb ik mij vier beperkingen opgelegd:

1. Als arbitraire grens is het jaartal 1945 genomen. Alleen publikaties vanaf dat jaar kwamen in aanmerking. Het doel van deze studie is niet een historisch onderzoek te verrichten naar oordelen over de taal van de overheid. Uit *Vervoorn* (1978) blijkt trouwens dat de kritiek van Van Lennep op de taal van de overheid in 1844 weinig verschilt van de hedendaagse kritiek. Wel is een paar keer een publikatie van vóór 1945 bij dit overzicht betrokken als ernaar werd verwezen in een publikatie van ná 1945 of als daarin een passage stond waarmee een bepaald verschijnsel goed kon worden geïllustreerd.

2. Voor zover het duidelijk aangegeven werd, zijn buiten beschouwing gelaten de oordelen over de taal in wetsteksten en het taalgebruik van politici. Een voorbeeld: het volgende citaat van *Lindner* in de Volkskrant van 1 november 1975 over het taalgebruik van politici is niet in het overzicht verwerkt:

"Maar de gespecialiseerde taal wordt doorweven met uitdrukkingen uit de administratieve en juridische sfeer, want er moeten nu eenmaal wetten en regelingen gemaakt worden. En dan komen er ook nog de speciale kreten bij van parlementaire traditie, hoffelijkheid, tactiek en politieke ideologie. Zo ontstaat een zeer eigenaardig taalkundig mengsel ..."

Dit houdt trouwens niet in, dat in dit citaat genoemde elementen niet voorkomen in het overzicht. In andere publikaties stonden namelijk ook over de taal van 'Den Haag' in het algemeen, opmerkingen over juridisch aandoende nauwkeurigheid en over hoffelijkheid. Wel treft u bijvoorbeeld in dit overzicht enkele opmerkingen aan van *Ritter* uit 1939 over de taal in wetten. (Naar hem verwijst onder anderen *Kamp* in 1961.) Ritter stelt namelijk dat zijn opmerkingen ook gelden voor de briefwisseling tussen overheidsdiensten.

3. Er is niet naar gestreefd alle publikaties te vermelden waarin opmerkingen over de taal van 'Den Haag' staan. Dat zou trouwens ook bijna onmogelijk zijn, gezien de grote hoeveelheid artikelen over dit onderwerp in kranten en personeelsbladen voor ambtenaren. Maar, afgezien daarvan, het zou ook niet nuttig zijn. In de meeste artikelen staan namelijk ongeveer dezelfde opmerkingen over de taal van 'Den Haag': Deze vorm van taalgebruik doet onpersoonlijk, ingewikkeld, langdradig en ouderwets aan. In teksten van de overheid staan veel dure woorden en ingewikkelde zinnen. Of zoals het in Elsevier van 10 juni 1972 staat: "... de nodeloze gewichtigdoenerij en de doolhof der zinsconstructies verhinderen elk begrip ..." En veel artikeltjes komen niet uit boven goed in het gehoor liggende scheldkanonnades, alle variërend op het thema 'ingewikkeldheid'.

Bij wijze van voorbeeld enkele opmerkingen uit het artikel van *Van Run* (1976), 'Wie dit leest is gek; wartaal en wantaal in nota's en rapporten': "Steeds dezelfde antieke inteelttaal. De gewone burger kan er niet bij en mag er blijkbaar niet bij ..." "Ze (hier zijn onder anderen ambtenaren en 'loslopende halfgeleerden' bedoeld, J.R.) willen alleen iets uitdrukken. Als zij hun formuleringen maar hebben afgescheiden op een zodanige manier dat alles erin staat en dat ze niet op leemtes of onjuistheden kunnen worden gepakt." "Op de taal van hun huis, wijk of buurt is een gammele bovenbouw gezet van boeken- of stenciltaal. En het

resultaat is een onpersoonlijk gestuntel met te moeilijke woorden. Wat hun ontbreekt is eenvoud, kritische zin en taalgevoel. Precies deze drie." Het zou weinig zinvol zijn al dergelijke, in verschillende toonaarden gegeven, algemene opmerkingen over de ingewikkeldheid van de overheids-taal te herhalen.

Deze derde beperking houdt ook in, dat in de paragrafen 3.3 en 3.4 bij de daar gesignaleerde kenmerken niet elke publikatie wordt genoemd, die aandacht besteedt aan zo'n kenmerk. Een voorbeeld: Bijna elke publikatie vermeldt het overmatig gebruik van de lijdende vorm als kenmerk van de taal van 'Den Haag'. Een opsomming zou nutteloos zijn; in zo'n geval zijn slechts de meest treffende citaten gegeven of citaten waarin ook nog een verklaring was te vinden voor het gesignaleerde verschijnsel.

4. Uiteraard kan niet gegarandeerd worden dat dit overzicht alle, waar ook maar gesignaleerde eigenschappen bevat, die aan de taal van 'Den Haag' worden toegeschreven. Het zoeken naar oordelen werd namelijk bemoeilijkt doordat deze vaak terloops worden gegeven. Zie bijvoorbeeld *Drop en De Vries* (1974:119,120) over de 'klemconstructie'. Vaak ook ontbreken in publikaties over verzorgd taalgebruik registers, onder andere in *Pieete* (1973). En *Hermkens* (1974:150) bijvoorbeeld geeft aan het eind van zijn boek in de paragraaf 'gecompliceerde zinsbouw' een opsomming van de "voornaamste kenmerken van de onpersoonlijke, geconcentreerde kanselarijstijl" terwijl het trefwoord 'kanselarijstijl' niet in het register voorkomt. Ook trof ik oordelen over overheidstaal aan in publikaties waarin men ze niet zou verwachten, bijvoorbeeld in *Japin* (1968) 'Taal en Techniek', terwijl daarentegen in de lezingen op het congres van het Genootschap Onze Taal van 1975 over 'Taal van Den Haag' geen enkel oordeel te vinden is. Maar na het voorjaar 1977, toen de materiaalverzameling voor deze studie werd afgesloten, zijn mij in het kader van mijn werkzaamheden daarna voor de Commissie Duidelijke Taal (tot 1979) geen wezenlijk andere oordelen ter ore gekomen.

Tot slot, in dit overzicht zijn alleen Nederlandse publikaties verwerkt. Maar ook in buitenlandse publikaties zijn soortgelijke opmerkingen te vinden als in dit hoofdstuk staan vermeld; zie bijvoorbeeld *Beauvais* (1970:115), *Ule* (1960:363-370), *Wagner* (1970:98), *Enkvist* (1964:22) en *Gowers* (1973: passim).

3.1.2 De opbouw van dit hoofdstuk

Na deze beperkingen bij het verzamelen van oordelen over het taalgebruik in teksten van de overheid, nu enkele opmerkingen over de opbouw van dit hoofdstuk.

Paragraaf 3.2 gaat over de aard van de oordelen. Aan de hand van enkele illustraties van algemene oordelen en gespecificeerde oordelen over zinsbouw- en woordkeusverschijnselen wordt aandacht besteed aan de vaagheid van de gehanteerde begrippen. Met algemene oordelen worden hier bedoeld de oordelen waarin geen taalkundige begrippen voorkomen. Bijvoorbeeld: "Het taalgebruik in teksten van de overheid is ingewikkeld en onnatuurlijk."

De eerste drie citaten in 3.2 zijn overigens de enige waarin eigenschappen van goed ambtelijk taalgebruik genoemd worden, of eisen die aan ambtelijk taalgebruik gesteld mogen worden. De kenmerken uit deze citaten konden teruggebracht worden tot vier positieve trefwoorden: zakelijk, bondig, duidelijk en waardig. Deze trefwoorden bleken een handig ordeningsprincipe op te leveren voor de paragrafen 3.3 en 3.4. De algemene negatieve oordelen worden apart behandeld in 3.3. In 3.3.2 worden deze oordelen gerubriceerd naar aanleiding van de vier trefwoorden hierboven. We zien dan dat deze algemene negatieve oordelen het beste zijn te karakteriseren met vier negatieve trefwoorden die inhouden dat de eigenschappen, aangeduid met de vier positieve trefwoorden, te ver zijn doorgevoerd, namelijk: onpersoonlijk, ingewikkeld, wijldlopij, traditioneel. Aan de hand van deze vier negatieve trefwoorden zijn de talrijke gespecificeerde opmerkingen over zinsbouw en woordkeus in teksten van de overheid gerangschikt in paragraaf 3.4, het belangrijkste onderdeel van dit hoofdstuk.

In paragraaf 3.4 staan dus de gespecificeerde oordelen waarin een min of meer taalkundig begrip wordt gebruikt, bijvoorbeeld: naamwoord, negatie, pleonasme. Deze oordelen zijn niet verder gesplitst naar opmerkingen over zinsbouw en woordkeus. Veel oordelen hadden namelijk betrekking op beide niveaus. Overigens valt paragraaf 3.4 voor een belangrijk deel te lezen als uitwerking van paragraaf 3.3. Voor het opstellen van hypotheses in hoofdstuk 6 is dan ook voornamelijk uitgegaan van de gespecificeerde oordelen genoemd in paragraaf 3.4. Tot slot volgen in paragraaf 3.5 nog enkele opmerkingen over de waarde van de oordelen voor het opstellen van hypotheses over de taal van 'Den Haag'.

3.2 De aard van de oordelen

Ter illustratie van de aard van de oordelen eerste enkele citaten. In 1966 verscheen de dissertatie van *Veering* 'Mogelijkheden en Moeilijkheden van Taalverzorging'. Deze dissertatie is in hoofdzaak gewijd aan het tijdschrift *Onze Taal*, een uitgave van het Genootschap Onze Taal. Dit genootschap heeft zich onder andere als doel gesteld: het bevorderen van een goed en doelmatig taalgebruik. *Veering* (1966:139) stelt het volgende:

"Over de zogenaamde ambtenarenstijl is al zoveel gezegd - ook op verschillende plaatsen in vele jaargangen van O.T. - dat er nauwelijks nog iets nieuws over te berde te brengen is. Men vergeet - ook bij een aantal zeker gerechtvaardigde grieven - al te gauw dat de taal van de overheid moet voldoen aan eisen die aan particulier taalgebruik niet altijd gesteld behoeven te worden: concentratie, onpersoonlijkheid, abstractie, zakelijkheid, waardigheid."

En in zijn artikel over het taalgebruik in de troonrede 'Knelpunten van procedurele aard' (1974:54) zegt *Veering* ongeveer hetzelfde als hij het heeft over de ambtelijke taal als een "juridische, administratieve vaktaal, gekenmerkt door onpersoonlijkheid, nauwkeurigheid, objectiviteit, ondubbelzinnigheid."

Het derde citaat komt uit het Belgische tijdschrift *Taalbeheersing in de Administratie*. Juist dit tijdschrift wordt hier aangehaald, omdat hierin heel overzichtelijk een aantal opmerkingen staan die ook telkens weer terug te vinden zijn in Nederlandse publikaties.

In de inleiding op een reeks artikeltjes over de 'Administratieve Stijl' (1962:I,11,12) staat de volgende belangrijke opmerking:

"De vormgeving, m.a.w. de stijl, kan bezwaarlijk worden losgedacht van de gegeven inhoud, en evenmin van de gegeven situatie (de verhouding van de steller tot de lezer)." Daarna wordt gesteld dat een goede administratieve stijl hoofdzakelijk wordt gekenmerkt door vier eigenschappen:

"Uiteraard zijn alle uitingen van de administratieve actie tot het verstand gericht. Daaruit volgt de hoofdvereiste van klaarheid. Zonder deze kan het beoogde doel niet bereikt worden.

Anderzijds is de bevrediging van het verstand niet het einddoel van de bestuursactie. Deze is op de uitvoering, op de daad afgestemd. De administratieve stijl wordt daarom ook door pragmatische bondigheid gekenmerkt.

De Administratie dient de gemeenschap, overeenkomstig de richtlijnen

van de overheid. De inhoud van de geschriften wordt niet eigenmachtig gecreëerd en de uiting is het produkt van samenwerking: daaruit stamt de onpersoonlijke objectiviteit van de ambtelijke stijl.

Ten slotte wordt de administratieve actie gevoerd in naam van de hoge overheid. Dit aspect verklaart de waardigheid die aan de administratieve redactie een eigen tonaliteit verleent."

Tot zover de citaten waarin alleen sprake is van algemene oordelen. Nu nog een voorbeeld van een gespecificeerd oordeel. De volgende opmerkingen over zinsbouw en woordkeus in ambtelijke teksten zijn ook afkomstig van *Veering* (1974:54):

"(De ambtelijke taal) is als vaktaal te vergelijken met de zakelijke wetenschapstaal. Enige opvallende trekken zijn de volgende:

1. De zin bevat veel gegevens. De geleiding in hoofd- en bijzinnen is ~~op~~ zichzelf meestal vrij eenvoudig. De bepalingen maken de zin zwaar; voorzetsels zijn frequent.
2. De eigenlijke informatie staat in het zelfstandig naamwoord; het werkwoord 'verbleekt' tot een soort hulpwerkwoord.
3. De zin is onpersoonlijk: voorkeur gaat uit naar passieve vormen ~~en~~ naar onbepaalde wijzen als zelfstandig naamwoord.
4. De taal is geconcentreerd: voorkeur voor één woord boven omschrijving."

Na deze citaten zal het duidelijk zijn, dat de oordelen waarin geen gebruik gemaakt wordt van taalkundige termen vaag blijven. En hiermee wordt niet bedoeld de vaagheid die voortvloeit uit het feit dat de oordelen impressionistisch zijn, dus niet op onderzoek gebaseerd, maar de vaagheid van de gehanteerde begrippen. In het derde citaat worden de volgende vier karakteristieken genoemd: klaarheid, bondigheid, objectiviteit en waardigheid. Maar wat is de precieze inhoud van deze termen? In de eerste twee citaten staan enkele, gedeeltelijk elkaar overlappende, kenmerken die veel gemeen hebben met de hierboven genoemde vier: concentratie, onpersoonlijkheid, abstractie, zakelijkheid, waardigheid, nauwkeurigheid, objectiviteit en ondubbelzinnigheid. Ook hier weer dezelfde vaagheid, een vaagheid die trouwens ook geldt voor alle algemene negatieve oordelen in paragraaf 3.3.

Het enige wat over de hier gegeven kenmerken gezegd kan worden, is dat er vier groepen zijn te onderscheiden:

1. abstractheid, objectiviteit, onpersoonlijkheid en zakelijkheid;
2. bondigheid en concentratie;
3. klaarheid, nauwkeurigheid en ondubbelzinnigheid;
4. waardigheid.

In het vervolg van dit hoofdstuk worden deze groepen aangeduid met de trefwoorden: 1. zakelijk; 2. bondig; 3. duidelijk; 4. waardig.

3.3 Algemene oordelen

In paragraaf 3.3.1 staan, in chronologische volgorde, citaten uit zes publikaties waarin, zonder dat er taalkundige termen worden gebruikt, gesproken wordt over de taal van 'Den Haag'. Door middel van deze citaten kan ik de belangrijkste algemene oordelen over het taalgebruik van de overheid de revue laten passeren. De oordelen worden in paragraaf 3.3.2 samengevat aan de hand van de vier hierboven gegeven trefwoorden.

3.3.1 Zes publikaties

In het boekje *Het orkest der overheid* (1939) maakt Ritter de volgende opmerkingen over de briefwisseling van de overheidsdiensten in het kader van de uitvoering of de voorbereiding van een wet:

"De grieven die tegen de taal onze wetten zijn in te brengen, gelden ook tegen de taal onze kanselarijen." Volgens Ritter ontbreekt het onze wetten - en dus ook de bedoelde briefwisseling - aan 'duidelijkheid' en 'onomwonden formuleering'. Dit komt omdat men er angstvallig naar streeft (a.w.:99,100) "uitzonderingen en uitzonderingen op uitzonderingen aanstonds in de hoofdbepaling te benoemen, in stede van ze samen te voegen in een afzonderlijke bepaling." Wel vertonen de ambtelijke bescheiden, meer dan de wetten, "een vooruitgang in logica van betoogtrant en bondigheid van zeggings." Maar er blijven veel fouten aan te wijzen. De grondfout is (a.w.:105) "dat de ambtelijke briefschrijver in den regel verzuimt de aandacht van zijn lezer te prikkelen. Vaagheid is de grootste zonde van den corresponderenden ambtenaar."

In *Een nieuwe ambtenaar, een nieuwe stijl* (1943) houdt Lans een pleidooi om de 'bureaustijl' te veranderen. Want: "In de beste gevallen doet de ambtenarenstijl alleen maar gewrongen en onnatuurlijk aan" (a.w.:4). En bovendien worden ambtelijke brieven gekenmerkt door onduidelijkheid.

Van Braam geeft in zijn dissertatie *Ambtenaren en bureaukratie in Nederland* (1957) een uitgebreide beschrijving van het ambtenarenapparaat en besteedt ook anderhalve pagina aan 'Ambtelijk en verambtelijk taalgebruik' (a.w.:324,325). Hij stelt dat "de ambtelijke taal het menselijk-pleizierige en heldere vaak zo sterk mist. De ambtelijke taal is dikwijls strak en afgemeten, befehlens, zakelijk-koel,

technisch en traditioneel." De wijldlopigheid viert hoogtij. En verder dragen ambtelijke brieven "duidelijk de kenmerken van specialistenjargon. Zij zijn bovendien uitdrukking van een, door vrees voor konsekwenties en het streven naar de hoogst mogelijke rechtvaardigheid en objectiviteit bepaalde, hang naar het volledige en het verantwoorde."

Persoonlijke onpersoonlijkheid was de veelzeggende titel van een rede die Van Poelje in 1960 hield op de jaarvergadering van de Maatschappij der Nederlandse Letterkunde. Een rede die de aandacht nog altijd waard is volgens *In 't Veld* (1977:13).

"Vraagt men, wat in het 'gewone' ambtelijke taalgebruik het meest opvalt, dan moet, naar mijn mening, het antwoord luiden, zijn onpersoonlijkheid." (a.w.:12) Deze onpersoonlijkheid, aldus Van Poelje, is noodzakelijk, want: "De ambtelijke brief moet onder haast alle omstandigheden een ander karakter dragen dan de particuliere correspondentie." (a.w.:14) Verder zijn de volgende twee opmerkingen nog van belang. "Buiten het gebied van de wetgeving kenmerkt het ambtelijk taalgebruik zich in zijn ontwikkeling over het algemeen door een grote versobering. Alle overdaad, alle lege frasen, alle bombast zijn geheel of voor het overgrote deel verdwenen." (a.w.:15) Maar Van Poelje noemt in dit verband alleen de titulatuur als voorbeeld en zegt daarna: "Afgezien van deze versobering heeft de ambtelijke taal over het algemeen een statisch, lichtelijk archaisch karakter."

In hetzelfde jaar, 1960, hield het Genootschap Onze Taal een congres over *De taal in het openbare leven*. In zijn rede over het ambtelijk taalgebruik haalt Kamp met instemming de volgende omschrijving aan: "... stadhuistaal met haar droge en saaie uitdrukkingen en bloedarme sjablone-achtige gedachtenuitingen ..." (Zie *Onze Taal* 1961:9.) En een pagina verder wordt ook nog genoemd: "... de vereniging van de juridische en de stijlkundige precisie: het zó te zeggen dat de ambtenaar aan de andere schrijftafel het dadelijk begrijpt en het zonder veel hoofdbreken kan uitvoeren."

Tot besluit nog twee citaten uit een artikel van *Wijngaarden* in de bundel *Macht over Macht* (1975) Wijngaarden zegt onder het hoofdje 'De problematiek van de ondoorzichtigheid' het volgende naar aanleiding van de taal van de overheid (a.w.:53):

"Gezagserkenning is niet mogelijk zonder gezagsherkenning. Die herkenning is niet mogelijk als er geen begrijpelijke taal gesproken wordt." Daarom moeten we, aldus Wijngaarden, oppassen met het gebruik van 'codetaal'. Codetaal is wel efficiënt tussen vakgenoten, maar wanneer

"in een democratische samenleving de 'bevoegde instanties' een onverstaanbare codetaal spreken tegen mensen wie hun beslissingen aangaan, werken zij hun eigen verstaanbaarheid en daarmee de erkenning van hun eigen gezag tegen."

3.3.2 Samenvatting

De vier positieve kenmerken van de taal van 'Den Haag', genoemd aan het einde van paragraaf 3.3.1 (zakelijk, bondig, duidelijk en waardig), geven een aardig handvat om alle, gedeeltelijk verschillende en gedeeltelijk overeenkomstige, negatieve algemene oordelen over het taalgebruik van de overheid te rubriceren. Uiteraard geldt ook voor deze indeling de opmerking uit paragraaf 3.2 dat de termen vaag zijn en voor een, moeilijk nader te definiëren, deel tot elkaars synoniemen gerekend kunnen worden. Het oordeel 'onnatuurlijk taalgebruik' bijvoorbeeld kan bij alle vier kenmerken worden ondergebracht. De hier gegeven indeling is alleen bedoeld om een beter overzicht te krijgen.

De algemene negatieve oordelen kunnen gerangschikt worden onder een van de vier genoemde kenmerken met de toevoeging 'te' (te ver doorgevoerd of overdreven):

1. te zakelijk - afgemeten, befehlerisch, droog, een hang naar objectiviteit en het verantwoorde, niet menselijk-plezierig, saai, strak, technisch, zakelijk-koel, onpersoonlijk enz.
2. te bondig - codetaal, gewrongen, niet helder, specialistenjargon, zoveel mogelijk gegevens in een zin, ingewikkeld enz.
3. te duidelijk - een hang naar het volledige, stijlkundige en juridische precisie, wijldlopig enz.
4. te waardig - archaïsch, statisch, traditioneel enz.

Deze groepen algemene negatieve oordelen duid ik in het vervolg aan met de trefwoorden: 1. onpersoonlijk; 2. ingewikkeld; 3. wijldlopig; 4. traditioneel.

Een heel aardige parallel is trouwens te trekken met de aanduidingen in woordenboekomschrijvingen onder 'kanselarijstijl' of 'kanselarijstaal' na de opmerking 'in ongunstige zin': een dorre (1), ingewikkelde (2), langdradige (3) en stijve (4) schrijftaal. Deze aanduidingen zijn te vinden in het *Woordenboek der Nederlandsche Taal*, deel VII, 1

(1926:1305), *Van Dale* (1976:1118), *Koenen-Endepols* (1977:618) en *Kramers* (1977:436).

Daarnaast geven deze woordenboeken nog als omschrijving: een schrijftrant doorspekt met vreemde of verouderde woorden of zinswendingen, formules en vaktermen. Deze verschijnselen worden in de volgende paragraaf behandeld. Daar komen overigens ook de publikaties uit 3.3.1 weer ter sprake, voor zover daarin iets wordt opgemerkt over zinsbouw en woordkeus.

3.4 Oordelen over zinsbouw en woordkeus

De grote verscheidenheid aan oordelen over zinsbouw en woordkeus heb ik zo systematisch mogelijk ondergebracht onder de vier trefwoorden, genoemd aan het einde van de vorige paragraaf:

1. onpersoonlijk; 2. ingewikkeld; 3. wijdlopig; 4. traditioneel.

Zoals reeds is opgemerkt, kunnen veel verschijnselen worden gerangschikt onder meer dan één trefwoord. Bijvoorbeeld, clichématig taalgebruik is niet alleen plechtstatig (traditioneel) maar vaak ook langdradig (wijdlopig). In de literatuur worden specifieke oordelen over zinsbouw en woordkeus vaak in verband gebracht met verschillende van de hierboven genoemde algemene oordelen. Bij de 'naamwoordstijl' bijvoorbeeld worden opmerkingen gemaakt over onpersoonlijk taalgebruik (1), het streven naar nauwkeuriger analysering (3) en waardig taalgebruik (4). En het gebruik van moeilijke woorden geeft niet alleen aanleiding tot de kwalificatie 'ingewikkeld' (2), maar ontlokt ook het oordeel 'nodeloze gewichtigdoenerij' (4). Terwille van de overzichtelijkheid zijn de oordelen over zinsbouw en woordkeus gerangschikt onder het algemene oordeel waarmee ze het meest in verband gebracht kunnen worden.

3.4.1 Onpersoonlijk

Onder dit eerste trefwoord vindt u opmerkingen over de naamwoordstijl (3.4.1.1), de lijdende vorm (3.4.1.2) en een aantal losse notities die zijn samengevat onder 'het menselijk element' (3.4.1.3).

3.4.1.1 Naamwoordstijl

Met naamwoordstijl wordt die vorm van taalgebruik bedoeld waarin (1) veel werkwoorden zijn gesubstantiveerd, vergelijk "Wij stellen ons primair als doel de steden te behouden en te herstellen" en "Onze

primaire doelstelling is het behoud en herstel van de steden" en
(2) werkwoorden zijn omschreven met behulp van een zelfstandig
naamwoord en een minder concreet werkwoord, bijvoorbeeld 'de opening
<verrichten' in plaats van 'openen'.

Hermkens (1974:91) merkt hierover het volgende op:

"Substantivering van werkwoorden is in het Nederlands normaal (...). De
Duitsers gaan echter verder dan wij. In ambtelijke, journalistieke en
wetenschappelijke taal wordt het Duits dikwijls ten onrechte nage-
bootst: daardoor wordt onze stijl zwaar en onpersoonlijk."

Veering (1969) signaleert dit verschijnsel in 'Het stervende werkwoord':

"Naamwoordelijke stijl is een van de hoofdkenmerken van zakelijke,
rationele stijl, zoals die bijvoorbeeld voorkomt in rapporten, ambte-
lijke verslagen, wetenschappelijke artikelen, alle bestemd voor een
bepaalde groep ter zake kundige lezers." (a.w.:54) Een paar voor-
beelden: 'tot ontwikkeling brengen' in plaats van 'ontwikkelen'; 'in
beschouwing nemen' voor 'beschouwen'. Deze manier van onder woorden
brengen is - aldus Veering - uitingsvorm van "een streven naar nauwkeuriger
analysering: de 'hulpwerkwoorden' geven schakeringen." Of anders gezegd:
"Het gebruik van de hulpwerkwoorden en de voorzetsels (...) heeft de
mogelijkheid tot een nauwkeuriger analyse van het aan te duiden feit
geschapen." En bovendien: "Dit type kan nog het voordeel hebben dat er
door de omschrijving een zekere waardigheid van de handeling tot uit-
drukking komt." (a.w.:55)

Ook Bakker (1964:15,16) signaleerde al armoede aan werkwoorden in het
'Officiaalees' en "het zgn. Expertinees, de meest gebruikte Rapportage-
taal, verwant aan het oude Ambtoe en het Burokraats." Zie hierover ook
het tijdschrift *Taalbeheersing in de Administratie* (1963:I,44), waarin
trouwens op verschillende plaatsen aandacht wordt besteed aan de
'substantiefstijl':

"Naast substantieven kunnen ook als zelfstandig naamwoord gebruikte on-
bepaalde wijzen de zin schragen, wat in nog hogere mate nefast (Dit
woord betekent ongeveer hetzelfde als 'funest'. J.R.) is voor een
vlotte stijl. Het wordt vooral erg wanneer de infinitief vergezeld is
van bepalende woorden." Een van de hier gegeven voorbeelden luidt:
"Ten einde het vervallen van hun eventueel recht op naturalisatie te
voorkomen."

Nauw verwant aan de naamwoordstijl is het veelvuldig gebruik van
woorden die eindigen op '-ing'. Zie Veering (1963a). Vergelijk de
volgende twee zinnen uit een voorbeeld over de 'ing-ziekte': "De beide
ministers hebben in hun gesprek kunnen constateren dat ze het eens zijn

over bijna alle kwesties ..." "De beide ministers zijn bij hun onderhandeling tot de constatering van hun in bijna alle kwesties aanwezige overeenstemming gekomen ..." "Nu is - aldus Veering (a.w.:39) - de zin de ingdood gestorven en daarmee is hij rijp voor het communiqué." Ook Ritter (1939:109) wees al op de 'ing-woorden' naar aanleiding van de monotonie der klankvorming in ambtelijke teksten.

3.4.1.2 Lijdende vorm

In veel publikaties wordt naar aanleiding van het taalgebruik van de overheid een opmerking gemaakt over het gebruik van de lijdende vorm. Zie bijvoorbeeld Japin (1968:43): "Meestal doet de passieve vorm ambtelijk-stijf aan naast de actieve vorm, die veel soepeler en natuurlijker is." Zie ook Kolkhuis Tanke (1975:51): "Een van de oorzaken waardoor vooral vele ambtelijke verklaringen (...) zo weinig aanspreken, is het onaflaatbare gebruik van de lijdende vorm." En: "De lijdende vorm doet bureaucratisch aan. Hij maakt, bij veelvuldig gebruik, de tekst onpersoonlijk, vaag en gewichtig."

In sommige publikaties wordt tevens een verklaring gegeven voor het gebruik van de lijdende vorm. Zo geeft Van Braam (1957:326) de volgende toelichting:

"Het veelvuldig gebruik van de lijdende vorm is (...) een symptoom van de vlucht voor de persoonlijke verantwoordelijkheid en beslissing, welke men zo dikwijls onder ambtenaren aantreft. Heines - Van Braam haalt hier een artikel aan uit het personeelsblad van het Ministerie van Financiën (1947:5) - formuleert het aldus: Het zelfgevoel, de persoonlijkheid wordt door die lijdende vorm min of meer opzij gezet, de ex professo schrijvende ambtenaar voelt zich als een deel van de ambtenarenwereld, als een deel van het onpersoonlijk gezag."

Soortgelijke opmerkingen plaatst Pleete (1973:89) onder het hoofdje 'passivitis'. En Hermkens (1974:134) zegt het weer anders:

"(Het passivum) is onpersoonlijker, algemener, bescheidener dan het activum. In zakelijke of ambtelijke stukken hanteert de auteur de lijdende vorm om op de achtergrond te blijven. De passieve constructie brengt distantie teweeg."

3.4.1.3 Het menselijk element

De notities onder dit hoofdje sluiten nauw aan bij de opmerkingen over de vlucht voor de persoonlijke verantwoordelijkheid naar aanleiding van het veelvuldig gebruik van de lijdende vorm en bij opmerkingen over het mogelijk onpersoonlijke karakter van de naamwoordstijl. In *Taalbeheersing in de Administratie* (1963:I,28) wordt een passage gewijd aan de tendens om in ambtelijke teksten de directe werkwoordsvormen te vervangen door afgezwakte uitdrukkingen waarin het accent op het naamwoord ligt:

"Het zijn precies de werkwoorden die een oordeel of een beslissing weergeven, welke het sterkst die tendens ondervinden, waardoor de indruk wordt gewekt dat de ambtenaar terugschrikt voor zijn verantwoordelijkheid. Deze indruk gaat nog sterker uit van de onpersoonlijke weergave van een oordeel of een beslissing, zoals dit het geval is in uitdrukkingen als: er wordt voorgesteld - het besluit dringt zich op (...). Ook het stelselmatig gebruik van de conditionalis en van het werkwoord 'schijnen' gaat dezelfde richting uit."

In dit verband moet ook vermeld worden de opmerking van *Hermkens* (1974:150) over "het gebruik van de 3e persoon (ondergetekende)."

Hermkens (a.w.: 120) brengt trouwens ook het gebruik van de niet-conjuncte, beknopte bijzinnen in verband met "een (ongemotiveerde) angst (...) voor het voornaamwoord ik". Hierover meer in 3.4.2.1, 'Bijzinnen en tussenzinnen'.

Tot slot, naar aanleiding van het onpersoonlijke karakter van de taal van 'Den Haag', een paar opmerkingen over eigenaardigheden die in studies over leesbaarheid meestal onder 'human interest' worden vermeld. In onderzoeken naar leesbaarheid van teksten wordt vaak gesteld dat, naast vraag- en uitroepzinnen, de directe aanspreekvormen de afstand tussen auteur en lezer verkleinen en de betrokkenheid bij de tekst verhogen. (Zie *Van Hauwermeiren*, 1975:151-155.)

Het tegenovergestelde is het geval bij een overmatig gebruik van het woordje 'men'. "In formele briefwisseling, in wetenschappelijke, ambtelijke of kerkelijke taal is 'men' onmisbaar" (*Hermkens*, 1974:57).

"Het nadeel van 'men' is, dat bij veelvuldig gebruik hiervan de stijl iets kanselarij-achtigs om niet te zeggen iets loodzwaars gaat krijgen" (*Veering*, 1971:119,120). *Hamer* (1975:19) signaleert het schaarse gebruik van de u-vorm in ambtelijke stukken, en het gebruik van de mededeling dat er een vraag gesteld wordt in plaats van de directe vraag. Bijvoorbeeld: "Wij verzoeken u ons een en ander spoedig te doen toekomen" in plaats

van "Wilt u ons een en ander spoedig zenden?" Vergelijk ook nog deze opmerking in *Taalbeheersing in de Administratie* (1963:I,33):

"In de administratieve stijl worden liefst geen vraagzinnen en zeker geen uitroepzinnen gebruikt." Deze opmerking komt na de vaststelling: "Elke affectieve geladenheid is trouwens misplaatst bij de administratieve redactie."

3.4.2 Ingewikkeld

Na de opmerkingen over het onpersoonlijk karakter van de taal van 'Den Haag' nu de opmerkingen die te maken hebben met ingewikkeld taalgebruik (het tweede trefwoord). Twee verschijnselen vergen een iets uitgebreidere behandeling: bijzinnen en tussenzinnen (3.4.2.1) en tangconstructies (3.4.2.2). Daarna volgen opmerkingen over 'elliptische constructies' en 'onhandige ontkenningen' in 3.4.2.3 onder 'andere zinsbouwverschijnselen'. En tot slot in 3.4.2.4 onder 'moeilijke woorden' de opmerkingen over woordkeus.

3.4.2.1 Bijzinnen en tussenzinnen

"Een opeenstapeling van toe- en tussenvoegsels, verwijzingen, voorbehouden en uitzonderingen dient men te vermijden." Dit advies uit de 'Aanwijzingen inzake indeling en stijl van ambtelijke stukken', vastgesteld bij besluit van de minister-president van 4 september 1970, is zeker niet overbodig gezien de talrijke opmerkingen over lange en ingewikkelde zinnen met veel bijzinnen en tussenzinnen.

Ritter (1939:108,109) spreekt hier al over. Hij wijst in dit verband ook op "het nimmer ophoudende woordje 'dat'", en op de overmaat aan komma's. Meyers (1965:82,83) wijst op:

"een hinderlijke voorkeur voor een ingewikkelde zinsbouw, waarbij de lezer de indruk krijgt, in een rangeertreintje te zitten, langs wissels en zijsporen voorthobbend, af en toe even op het hoofdspoor terechtkomend om dan weer langs andere wissels en zijsporen voort te sukkelen. Tot dan eindelijk bij de punt achter de zin het einde is bereikt, waar hij even kan uitblazen." In het tijdschrift *Taalbeheersing in de Administratie* (1967:I,160,176) wordt deze manier van zinsbouw aangeduid met 'lintwormstijl' en 'kettingzinnen'. En verder wordt daar nog gewaarschuwd tegen "het vooropzetten van verscheidene bepalingen".

De opmerkingen na 1970 over de zinsbouw, bewijzen dat de 'Aanwijzingen' van de minister-president op dit punt - en trouwens ook op andere -

weinig effect hebben gehad. *Pieete* (1973:75) zegt naar aanleiding van de verbindingen zoals 'ten bewijze waarvan', 'in verband waarmee':

"Deze verbinding is kenmerkend voor de ambtelijke stijl, die ernaar streeft datgene wat meegedeeld moet worden, in één zin samen te persen, gaat het niet goedschiks, dan maar kwaadschiks." *Hermkens* (1974:150) noemt in dit verband als kenmerken de beknopte bijzin en de tussenzin. *Veering* (1974:54) daarentegen laat een ander geluid horen: "De geleding in hoofd- en bijzinnen is op zichzelf meestal vrij eenvoudig. De bepalingen maken de zin zwaar ..." Verder zijn opmerkingen over ingewikkelde zinsbouw terug te vinden bij *Hamer* (1975:10-13) en in *Moerdijk en Sluimer* (1976:180).

3.4.2.2 Tangconstructies

Naast een overmaat aan bij- en tussenzinnen wordt ook vaak een ander verschijnsel genoemd als het gaat om ingewikkeld taalgebruik, namelijk een al te grote afstand tussen zinsdelen die bij elkaar horen. Vergelijk in het volgende voorbeeld uit het artikel 'Spanning in de zin' van *Veering* (1963b:64) de afstand tussen 'de' en 'verandering' en tussen 'verandering' en 'kan ... verbeteren': "De door de Tweede Kamer beoogde verandering in de procedure van de begrotingsbehandeling, waartoe dit jaar wederom een proef zal worden genomen, kan mettertijd het overleg over deze keuze verbeteren."

In *Rogier-Pieete* (1954:36) wordt dit verschijnsel als volgt onder woorden gebracht: "Sommige ambtelijke stilisten construeren hun perioden zo, dat geen enkele lezer de zin er van zonder teruglezen en napluizen kan vatten. Hun ideaal is de synthese, de samenvatting van de geordende veelheid in de eenheid van één enkel zinsverband." Als kenmerk wordt hierbij vermeld dat tussen lidwoord en zelfstandig naamwoord staat wat 'een gewoon stilist' als bepaling of bijvoeglijke bijzin achter het zelfstandig naamwoord zet.

De 'spanning in de zin' blijft natuurlijk niet alleen beperkt tot die tussen lidwoord en zelfstandig naamwoord. In *Taalbeheersing in de Administratie* (1967:I,176) worden ook nog als voorbeeld aangehaald de afstand tussen onderwerp en werkwoord, de afstand tussen werkwoord en lijdend voorwerp, en impliciet - via het waarschuwen voor 'het samen-treffen van voegwoorden' - ook de spanning die ontstaat door de bijzin in de bijzin. (Zinnen van het type: "Het is onvermijdelijk dat, omdat ik veel moet citeren, dit hoofdstuk langdradig wordt.")

Ook *Pieete* (1973:103,104) wijst naar aanleiding van "het maken van lange tussenzinnen en het invlechten van steeds weer nieuwe bijzinnen in de hoofdzin" op dit verschijnsel, waarvoor hij de benaming 'omsluiting' gebruikt. De benaming 'klemconstructie' wordt gebruikt in *Drop en De Vries* (1974:119,120). Hiermee wordt in het bijzonder bedoeld op de beknopte bijzin tussen lidwoord en zelfstandig naamwoord, bijvoorbeeld "de mij door verscheidene instanties gegeven adviezen".

Drop en De Vries zeggen hierover het volgende:

"Een echt akelig formele stijl krijgt u door een bepaalde combinatie van samentrekken (beknopt maken) en in elkaar schuiven die wel klemconstructie wordt genoemd." En na een voorbeeld volgt de uitspraak:

"U hoeft maar een paar van zulke constructies te gebruiken om uw stijl een formele, ambtelijke tint te geven - aangenomen natuurlijk dat het effect niet vloekt met de rest van uw stijl." Overigens is deze uitspraak niet meer terug te vinden in de tweede druk van deze publikatie (*Drop en De Vries*, 1977), waar over tangconstructies wordt gesproken op pagina 222 en 223.

Een algemene benaming voor het verschijnsel dat er een grote afstand zit tussen bij elkaar behorende zinsdelen, is 'tangconstructie' of 'wigconstructie'. Zie hierover *Hermkens* (1974:131):

"In tegenstelling tot het Frans heeft onze taal een 'onlogische' woordvolgorde: elementen die semantisch het nauwst verbonden zijn, worden syntactisch gescheiden; andere zinsdelen dringen er als een wig tussen, ze worden erdoor omsloten als door de grijpers van een tang." Op pagina 132 vervolgt *Hermkens*: "Een bijzonder type tangconstructie treffen we vooral aan in ambtelijke kringen: de lange bepaling tussen lidwoord (of voornaamwoord) en substantief (...). Meestal volgt onmiddellijk op het lidwoord een voorzetsel." Vergelijk zinswendingen als "De in de door mij dezer dagen verleende machtiging genoemde voorwaarde" en "Het maximum ... bedraagt 70% van de voor de ambtenaar voordeligste pensioengrondslag", constructies die aan *Lans* (1943) al de kwalificatie 'ambtelijk stotteren' ontlokten. Vergelijk ook de 'Van-de-voor-de-variété' van *Milot* (1957).

3.4.2.3 Andere zinsbouwverschijnselen

Een ander verschijnsel in teksten van de overheid waarover opmerkingen zijn gemaakt, is de elliptische constructie. *Hermkens* (1974:89,150) spreekt hierover bij de behandeling van de constructie "Dat is te maken" waarbij een noodzakelijkheid wordt uitgedrukt. "In ambtelijke taal is algemeen gebruikelijk: de te nemen maatregelen, de te bespreken punten ...". "Een andere ambtelijke constructie is die zonder persoonsvorm:

'Dit formulier geheel met inkt in te vullen'." Om deze laatste constructie gaat het hier. Ook *Lans* (1943) noemde al het verschijnsel van de 'ellipsomanie' met voorbeelden als "Het bedrag dient gestort". Verder spreekt *Ritter* (1939:109), en in navolging van hem *Kamp* (1961:10), over onhandige ontkenningen. Hoewel deze auteurs geen voorbeelden geven, blijkt uit het verband dat we moeten denken aan constructies als "Vermeden moet worden dat iets niet gebeurt" en "Het is niet onaannemelijk dat". Vergelijk ook *Hamer* (1975:18,19) over de 'negatieve stijl': "In ambtelijke kringen schijnt een zekere schroom te bestaan om mensen direct te benaderen. De lezer komt te weten hoe iets niet is." Een voorbeeld: "Alhoewel aan de bestemming 'opslagplaats' abusievelijk geen goedkeuring is onthouden."

3.4.2.4 Moeilijke woorden

Hieronder rangschik ik de opmerkingen over afkortingen, vreemde woorden en vakjargon.

Vrij algemeen is de klacht dat ambtelijke teksten ingewikkeld worden, doordat allerlei instellingen en commissies aangeduid worden met afkortingen. Of zoals *Van Poelje* (1957:214) het zegt: "(Er bestaat een streven) de wereld ... nog moeilijker verstaanbaar te maken door allerlei instellingen en begrippen niet aan te duiden met naam of omschrijving, maar met een combinatie van losse letters."

En verder wordt in bijna elk kranteartikel over de taal van de overheid het specialistenjargon of de overmaat aan vaktermen aan de kaak gesteld. Vooral het CRM-jargon of veldwerkersjargon moet het ontgelden. *Hermkens* (1974:150) signaleert in dit verband ook het gebruik van 'bastardwoorden' als 'trend'. *Van Braam* (1957:324) noemde het specialistenjargon een belemmering voor de begrijpelijkheid. En in 3.3.1 is al vermeld hoe *Wijngaarden* (1975:53) sprak over de codetaal tussen vakgenoten in stukken bedoeld voor niet-deskundigen.

3.4.3 Wijdlopig

Na de opmerkingen over onpersoonlijk (3.4.1) en ingewikkeld (3.4.2) taalgebruik nu de opmerkingen die te maken hebben met het derde trefwoord, wijdlopig.

Ritter (1939:108) stelde de omhaal van woorden al aan de kaak:

"... vaagheid ontstaat in de eerste plaats door noodeloze 'verbredingen'." Verbredingen ontstaan door a. 'dooie woorden' ('de genoemde

Minister', 'de voormelde aangelegenheid' enz.); b. 'verwijzingen naar vorige bescheiden' of c. 'te uitvoerige vermeldingen van wettelijke bepalingen'. Vergelijk in dit verband ook *Lans* (1943) die als oorzaak van onduidelijkheid het voorkomen van 'overbodig referentiemateriaal' noemt.

Nauw verwant aan de zogenaamde dode woorden is het pleonastisch taalgebruik: "Pleonasme komt niet alleen voor bij 'slordige schrijvers', maar ook bij Jantje Sekuur, een verre neef van Tante Betje. Let maar eens op overduidelijke ambtelijke stijl" (*De Vooy* 1941:38). Ook in het tijdschrift *Taalbeheersing in de Administratie* wordt het pleonasme genoemd in het kader van de 'woordenkramerie' (1967:I,176,178) en de 'langdradigheid' en 'zinledige en overvloedige woordenvloed' die de stijl 'verwateren' (1967:I,161). In hetzelfde tijdschrift (1963:I,26-28) werd ook al gesproken over de toevoeging van overbodige bepalingen of 'stijlparasieten', die de essentiële woorden van hun draagkracht ontdoen. "De beste schuilhoeken voor stijlparasieten zijn de bijvoeglijke naamwoorden en de bijwoorden." ('Nogmaals herhalen', 'een strikt verbod' enz.)

Moerdijk en Sluimer (1976:182-186) besteden eveneens aandacht aan het verschijnsel 'Woorden te veel'. Hier worden onder andere de volgende verschijnselen gesignaleerd: pleonasme, tautologie, overbodig gebruik van: kunnen, zullen, mogen enz. en het onnodig gebruik van 'om'. Verder signaleren *Moerdijk en Sluimer* (a.w.:17) de herhaling:

"We treffen in ambtelijke stukken wel eens zinnen aan die - zij het ook in andere woorden - herhalen wat in een voorafgaande zin beweerd is."

3.4 (4) Traditioneel

Tot slot de opmerkingen over het taalgebruik in teksten van de overheid, die te maken hebben met het vierde trefwoord, traditioneel. Het betreft hier enkele opmerkingen over het gebruik van archaïsmen, germanismen die stijf of ouderwets aandoen en de aanvoegende wijs, gerangschikt onder 'boekentaal' (3.4.4.1), en het gebruik van clichés en overdreven beleefde uitdrukkingvormen onder 'clichématig taalgebruik' (3.4.4.2).

3.4.4.1 Boekentaal

Van de talrijke opmerkingen over archaïsch taalgebruik, deftige taal of boekentaal in teksten van de overheid volgen hier een paar voorbeelden: *Hermkens* (1974:150) vermeldt "archaïsmen als hetgeen, wier, hetwelk, moge" en "plechtige woorden als waar, uiteraard, onderhavig". Onder 'Boekenwoorden' (a.w.:57) zegt hij:

"In formele of verheven stijl is het betr. vnw. welk(e) evenmin uitgesloten. Maar in ambtelijke stukken en zelfs in kranteartikelen wordt het zo kwistig en zo nodeloos gebruikt, dat de verdenking van gewichtigdoenerij in ons opkomt." Over het woordje 'waar' als voegwoordelijk bijwoord van oorzaak, reden of tegenstelling zegt *Hermkens* (a.w.:48): "Deze aanwending van waar behoort tot de precieuze (aanstellerige) stijl. Ambtelijke stukken, kranteartikelen en vooral publikaties uit de studentenwereld worden erdoor ontsierd." (Bijvoorbeeld "Waar DOS tegen Blauw-Wit in gebreke bleef, staat het niet bovenaan op de ranglijst.")

Ook *Moerdijk en Sluimer* (1976:161) noemen de archaïsmen als kenmerk van het ambtelijk taalgebruik: "Het publiek vindt de ambtelijke stijl dan ook dikwijls stijf en niet vrij van archaïsmen." Enkele pagina's verder wordt - overigens zonder dat hierbij het kenmerk 'verouderd' genoemd wordt - gesproken over "Het typisch ambtelijk gebruik van de betrekkelijke voornaamwoorden welk en welke ..." (a.w.:171) en over 'hetgeen als betrekkelijk voornaamwoord' dat in de ambtelijke stijl 'ook nog wel eens' voorkomt in plaats van 'dat' (a.w.:172). Op dezelfde pagina nog een opmerking over 'hetwelk': "Ook het gebruik van het betrekkelijk voornaamwoord hetwelk in plaats van dat is ouderwets en overdreven plechtig." Vergelijk in dit verband ook *Meyers* (1965:84): "De kanselarijtaal zet graag de eenvoudigste vormen dik aan." Als voorbeelden noemt hij onder andere: het gebruik van de drie woorden 'alsmede', 'mitsgaders' en 'alsook' voor 'en', en 'hervatten van werkzaamheden' voor 'weer aan het werk gaan'. *Hermkens* (a.w.:150) noemt nog een ander kenmerk, namelijk het gebruik van "germanistische woorden als middels". En *Pieete* (1973:54) vermeldt onder het hoofdje 'Germanismen', na de opmerking dat het Nederlands het woord 'betreffende' alleen als voorzetsel kent: "Vooral in ambtelijke stijl wordt 'betreffende' echter ook als bijvoeglijk naamwoord gebruikt." Verder geeft *Pieete* (a.w.:71) nog een oordeel dat in deze paragraaf past, namelijk "Het gebruik van wat ... betreft berust ... vaak op deftigdoenerij."

Ook het gebruik van de 'archaïstisch aandoende' (*Moerdijk en Sluimer*, 1976:31) aanvoegende wijs hoort onder het hoofdje 'Traditioneel' thuis. Vergelijk bijvoorbeeld *Hamer* (1975:15): "Een versleten werkwoordsvorm is de aanvoegende wijs. Ze beklijft nog (...) in ambtelijke stukken."

3.4.4.2 Clichématig taalgebruik

Na een reeks voorbeelden van oude woorden en omslachtige uitdrukkingen, waaronder 'teneinde', 'doen toekomen' en 'ondanks de hiervoor geschetste omstandigheden' zegt *Hamer* (1975:15):

"Veel van deze termen doen niet alleen omslachtig, maar ook plechtig aan. In ambtelijke kringen vindt men ze veelal 'gewoon'. Ze zijn al te gewoon, ze zijn cliché's geworden (...). Het clichématige is het kenmerk van ambtelijk taalgebruik. Het is niet alleen een kwestie van wat verouderde termen en zinswendingen, het kleeft aan de gehele stijl." *Van Braam* (1957:324) haalt nog een schrijver aan uit het personeelsblad van het Ministerie van Financiën, die spreekt van een "algemeen gebezigde, doch beschimmelde clichébrievenstijl". En *Milot* (1953) schrijft: "De voornaamste stijlfout is het gebruik van de z.g. cliché's, een aantal staande uitdrukkingen, die door een niet oordeelkundig gebruik volkomen ontmunt zijn."

Onder dit hoofdje rangschik ik ook de opmerkingen van *Moerdijk en Sluimer* (1976:185,186) over 'overdreven beleefdheid' en 'portaalvrees' omdat hier alleen voorbeelden gegeven worden van zogenaamde standaarduitdrukkingen. "Vaak - aldus *Moerdijk en Sluimer* over ambtelijke correspondentie - zien we dat wellevendheid tot in het overdrevene wordt beoefend." Als voorbeelden geven zij onder andere "Uw zeer gewaardeerd schrijven" en "Wij delen u beleefd mede". Verder noemen zij als voorbeelden van de angst om met de deur in huis te vallen "Graag wil ik het volgende onder uw aandacht brengen" en "Van deze gelegenheid maak ik gaarne gebruik". Vergelijk tot slot ook de 'Aanwijzingen voor het schrijven van ambtelijke stukken' in het *Vademecum van het Ministerie van Onderwijs en Wetenschappen* (1976:87). Na een pleidooi voor hoffelijk taalgebruik volgt de opmerking: "... hoffelijkheid is niet het gebruik van allerlei verouderde schrijftaaltvormen zonder enige betekenis, zoals 'Het zij mij voorts vergund het navolgende onder Uw aandacht te brengen'."

3.5 De waarde van de oordelen

De waarde van de in dit hoofdstuk gegeven oordelen over de taal van 'Den Haag' wordt mijns inziens beperkt door de volgende negen factoren:

1. De houding van de lezer

Bij de oordelen over de taal van de overheid valt het moeilijk uit te maken, of de oordelen niet zijn ingegeven door de houding tegenover het overheidsapparaat in plaats van door eigenaardigheden in overheids-taal. Het kan best voorkomen dat een lezer gefirriteerd is door het onoverzichtelijke en ingewikkelde ambtenarenapparaat, waarin hij door langzaam malende molens van het kastje naar de muur wordt gestuurd, en dat de lezer zijn irritatie uit in kritiek op de taal van 'Den Haag', die dan maar al te gauw de kans loopt 'onoverzichtelijk' en 'ingewikkeld' genoemd te worden. Ook interesse en leeservaring spelen hier een belangrijke rol. En het is zelfs niet ondenkbaar dat een tekst van de overheid zonder meer als moeilijk wordt beoordeeld, omdat de overheid nu eenmaal de naam heeft moeilijke teksten te produceren.

2. Inhoud en vorm

Strikt genomen heeft een opmerking als "de zin bevat veel gegevens" ook te maken met de inhoud. Uiteraard kan de inhoud niet geheel buiten beschouwing blijven bij opmerkingen over vorm of taalgebruik, maar de kans bestaat wel dat bijvoorbeeld het inhoudsoordeel 'technische informatie' geuit wordt in het vorm-oordeel 'ingewikkelde zinsbouw'. Terwijl het best zo kan zijn dat die technische, voor de lezer moeilijke, informatie in eenvoudig gebouwde zinnen is gepresenteerd.

3. Vorm en inhoud

Ook is het niet ondenkbaar dat alleen al de kop van een ambtelijke brief met de postkamer- en archiefaanduidingen de lezer een oordeel ontlokt over de hele brief ("Weer zo'n ingewikkelde ambtelijke brief!"), of dat alleen het ontbreken van de aanhef tot het oordeel "een onpersoonlijke brief" voert, of zelfs dat het gebruik van de schreefloze letter (een letter zonder horizontale streepjes, vergelijk M en M) leidt tot het oordeel "saai en kleurloos taalgebruik."

4. De taal van 'Den Haag' en ander taalgebruik

In *Taalbeheersing in de Administratie* (1962:I,13) wordt na het in 3.2 gegeven citaat met de kenmerken 'klaarheid, bondigheid, objectiviteit, waardigheid' gezegd:

"Het is duidelijk dat sommige van deze eigenschappen ook in andere genres worden aangetroffen. Doch daar zijn ze dan weer anders verbonden."

Het is dan ook de karakteristieke verbinding van de bovengenoemde kwaliteiten welke de administratieve stijl identificeert."

Een bewijs hiervoor lijkt mij moeilijk te leveren. Gelukkig stellen anderen zich iets voorzichtiger op als het gaat om exclusieve kenmerken van de taal van 'Den Haag'. Meestal wordt er geen duidelijk onderscheid gemaakt tussen het taalgebruik van de overheid en bijvoorbeeld het taalgebruik van wetenschapsmensen of deskundigen.

5. De taal van 'Den Haag'

Zodra deze term of een synoniem ervan wordt gebruikt, blijft het onduidelijk of de opmerkingen van toepassing zijn op politiek, juridisch of ambtelijk taalgebruik, op twee van deze drie of op alle drie tegelijk, terwijl het heel goed mogelijk is dat deze drie typen taalgebruik van elkaar verschillen.

6. Verschillende soorten overheidsteksten

Vaak is het onduidelijk op welke soorten teksten wordt gedoeld: brochures van voorlichtingsdiensten, formulieren, brieven aan burgers of interne discussiestukken enz. Bijna nergens wordt onderscheid gemaakt tussen de verschillende soorten overheidsteksten. Dus valt het, strikt genomen, meestal niet uit te maken of oordelen over de taal van 'Den Haag' eigenlijk alleen maar betrekking hebben op bijvoorbeeld het taalgebruik in de toelichting op het belastingformulier.

7. Extreme voorbeelden

Vooraf de taalrubrieken in personeelsbladen van departementen en gemeenten munten uit in het citeren van enkele heel lange en ingewikkelde zinnen doorspekt met vakjargon, zonder daarbij te vermelden hoe vaak dergelijke zinnen voorkomen in bepaalde overheidsteksten. Zo wordt de suggestie gewekt dat elke tekst van de overheid vol staat met lange, ingewikkelde zinnen en moeilijke woorden.

8. Elkaar tegensprekende oordelen

In paragraaf 3.2 kwam ter sprake dat de geleding in hoofd- en bijzinnen in ambtelijke teksten meestal vrij eenvoudig is. Maar in 3.4.1.2 kwamen we het oordeel tegen dat de taal van 'Den Haag' juist gekenmerkt wordt door veel bijzinnen. Evenzo kan er een tegenspraak zitten in de kenmerken 'klaarheid' en 'bondigheid' uit het eerste citaat in paragraaf 3.2 (daar overigens al aangegeven met het woord 'Anderzijds'). En in 3.3.1 staan nog de mogelijk tegenstrijdige oordelen 'wijdlopieg' en 'gewrongen'. Bij al deze oordelen valt niet uit te maken of er verschillende categorieën teksten in het geding zijn (zie factor 6).

9. Vage termen

Op pagina 29 is al gesproken over het feit dat de oordelen vaag blijven door de gehanteerde termen (onpersoonlijk, ingewikkeld, wijdlopig, traditioneel taalgebruik enz.). Uiteraard geldt deze opmerking niet voor de meer concrete oordelen over zinsbouw- en woordkeusverschijnselen. Maar naar aanleiding van de algemene oordelen zal de onderzoeker zelf moeten bepalen welke verschijnselen hij wil onderzoeken.

In deze studie is de keuze van de te onderzoeken verschijnselen bepaald door het feit dat een uitgebreid, algemeen analyseschema al op ruime schaal is toegepast om verschillende vormen van taalgebruik in kaart te brengen. Ik doel hier op het codeersysteem van de Werkgroep Frequentie-Onderzoek van het Nederlands (zie pagina 4). Het nu volgende hoofdstuk is gewijd aan dit codeersysteem en de voor deze studie relevante gegevens over woordfrequenties. Daarna worden in hoofdstuk 5 gegevens gepresenteerd over de met dit systeem gecodeerde steekproef uit de taal van 'Den Haag'. En in hoofdstuk 6 volgen dan, binnen het kader van het hier gekozen codeersysteem, de hypothesen waarmee de hierboven gegeven oordelen worden getoetst.

4.1 Inleiding

In deze studie is gebruik gemaakt van een codeersysteem dat bekend staat onder de naam C_3C (codering met drie cijfers). Dit systeem is ontworpen door de Werkgroep Frequentie-Onderzoek van het Nederlands en toegepast in *Woordfrequenties* (red. Uit den Boogaart, 1975) en *Spreektaal* (red. De Jong, 1979). Een korte toelichting op C_3C is te vinden in *Woordfrequenties* (pag. 17-23). Waarom de Werkgroep dit systeem heeft gekozen, staat vermeld in het niet-gepubliceerde 'Verslag van de Werkgroep' (pag. 92-99). En verder is een uitgebreide toelichting te vinden in de eveneens niet-gepubliceerde 'Voorschriften lexicale codering in het systeem C_3C ' (voortaan: het Codeervoorschrift).

De keuze voor C_3C in dit onderzoek lag voor de hand. Het is het enige systeem waarmee al een groot aantal teksten zijn gecodeerd. Op deze manier wordt een vergelijking veel gemakkelijker. Wel is nog overwogen om uit de talrijke mogelijkheden van C_3C alleen die te gebruiken, die van belang zijn voor het onderzoek naar eigenaardigheden in de taal van 'Den Haag'. Maar bij een eerste verkenning bleek dat een groot aantal coderingen in C_3C ook gebruikt zouden moeten worden voor dit onderzoek. Daarom is besloten het codeersysteem in z'n geheel toe te passen. Bovendien bestaat zo de mogelijkheid dat bij voortgezette studie (zie verder paragraaf 7.4) de voor dit onderzoek gecodeerde tekstfragmenten op allerlei andere, binnen het kader van C_3C te onderzoeken, verschijnselen vergeleken kunnen worden met teksten die op dezelfde manier gecodeerd zijn.

In *Woordfrequenties* (pag. 17) staat dat de daar gegeven toelichting over het algemeen voldoende is om de codecijfers te interpreteren. Maar voor een goed inzicht in het codeersysteem is het toch belangrijk nader in te gaan op het niet-gepubliceerde Codeervoorschrift (4.2). Want de woordsoorten genoemd in C_3C bevatten soms andere woorden dan men op het eerste gezicht zou vermoeden. Ook is deze bespreking nodig om straks, in 4.3, een aantal punten van kritiek op C_3C aan de orde te kunnen stellen. Paragraaf 4.4 bevat een overzicht van de frequentielijsten die in *Woordfrequenties* zijn gepubliceerd. En in 4.5 volgen tot slot nog enkele opmerkingen over de waarde van deze lijsten.

4.2 Het codeersysteem C_3C

C_3C is een lexicale codering die als doel heeft woorden te onderscheiden naar hun grammaticale functie en eventuele morfologische eigenschappen. Achter elk woord komt een code van drie cijfers. Soms komt er nog een vierde cijfer bij (zie pagina 68). En in een enkel geval geldt een spatie niet als woordgrens. Dit is bijvoorbeeld het geval bij eigennamen van meer dan één woord (Jan Jansen); deze eigennamen worden als één woord gecodeerd.

In de nu volgende toelichting wordt van de meeste coderingen die in *Woordfrequenties* voorkomen een voorbeeld gegeven. Bovendien worden steeds de minder voor de hand liggende beslissingen vermeld. Overigens blijkt uit de frequentielijsten van codes, de lijsten D_1 en D_2 , dat er van de 1000 mogelijke coderingen in C_3C slechts 190 voorkomen. Nog één opmerking vooraf: Een punt in een codering geeft aan dat op die plaats nog een cijfer uit het codeersysteem moet worden ingevuld, waarover elders uitleg wordt gegeven.

Nu eerst, in 4.2.1, een overzicht van C_3C ; daarna in 4.2.2 de toelichting. Op de terminologie kom ik terug in 4.3.1.

4.2.1 Overzicht C_3C

DE EERSTE TWEE CIJFERS

- 0.. als 1e cijfer: *substantieven*
 - 0. als 2e cijfer: 'gewone' substantieven
 - 1. " " " : eigennamen
 - 2. " " " : adjectivisch gebruikte substantieven
 - 8. " " " : interjectivisch gebruikte substantieven
 - 9. " " " : woorden in zelfnoemfunctie
 - 1.. als 1e cijfer: *adjectieven*
 - 0. als 2e cijfer: 'gewone' adjectieven
 - 2. " " " : substantivisch gebruikte adjectieven
 - 5. " " " : adverbiaal gebruikte adjectieven
 - 8. " " " : interjectivisch gebruikte adjectieven
- 2.. als 1e cijfer: *werkwoordsvormen*
 - deelwoorden en infinitieven
 - 0. als 2e cijfer: van intransitief gebruikte werkwoorden
 - 1. " " " : van transitief gebruikte werkwoorden
 - 2. " " " : van reflexief gebruikte werkwoorden
 - 3. " " " : van hulp- en koppelwerkwoorden
 - persoonsvormen
 - 4. " " " : van intransitief gebruikte werkwoorden
 - 5. " " " : van transitief gebruikte werkwoorden

- 6. " " " : van reflexief gebruikte werkwoorden
 - 7. " " " : van hulp- en koppelwerkwoorden
 - 8. " " " : van interjectivisch gebruikte werkwoorden
-
- 3.. als 1e cijfer: *pronomina A*
 - 0. als 2e cijfer: *personalia*
 - 2. " " " : *possessiva* (zelfstandig)
 - 3. " " " : *possessiva* (bijvoeglijk)
 - 4. " " " : *reflexiva* (zelfstandig)
 - 5. " " " : *reflexiva* (bijvoeglijk)
 - 6. " " " : *demonstrativa* (zelfstandig)
 - 7. " " " : *demonstrativa* (bijvoeglijk)
-
- 4.. als 1e cijfer: *pronomina B*
 - 0. als 2e cijfer: *interrogativa* (zelfstandig)
 - 1. " " " : *interrogativa* (bijvoeglijk)
 - 2. " " " : *relativa* (zelfstandig)
 - 3. " " " : *relativa* (bijvoeglijk)
 - 4. " " " : *indefinita* (zelfstandig)
 - 5. " " " : *indefinita* (bijvoeglijk)
 - 6. " " " : *cardinalia* (zelfstandig)
 - 7. " " " : *cardinalia* (bijvoeglijk)
 - 8. " " " : *ordinalia* (zelfstandig)
 - 9. " " " : *ordinalia* (bijvoeglijk)
-
- 5.. als 1e cijfer: *bijwoorden*
 - 0. als 2e cijfer: 'gewone' bijwoorden
 - 1. " " " : aanwijzende en onbepaalde bijwoorden
 - 2. " " " : vragende bijwoorden
 - 3. " " " : betrekkelijke bijwoorden
 - 4. " " " : voornaamwoordelijke bijwoorden
 - 5. " " " : vragende voornaamwoordelijke bijwoorden
 - 6. " " " : betrekkelijke voornaamwoordelijke bijwoorden
 - 8. " " " : interjectivisch gebruikte bijwoorden
-
- 6.. als 1e cijfer: *pre- en postposities*
 - 0. als 2e cijfer: *voorzetsels*
 - 1. " " " : 2e deel gesplitste voornaamwoordelijke bijwoorden
 - 2. " " " : niet-verbaal deel van samengestelde werkwoorden
 - 3. " " " : 2e deel gescheiden voorzetselgroepen
 - 4. " " " : achterzetsels
 - 5. " " " : te bij infinitief en voorzetsel met een te+infinitief-constructie als rest
 - 6. " " " : voorzetsels met een door onderschikkend voegwoord ingeluide bijzin als rest
-
- 7.. als 1e cijfer: *conjuncties*
 - 0. als 2e cijfer: *nevenschikkende voegwoorden*
 - 1. " " " : onderschikkende voegwoorden
 - 2. " " " : voegwoorden van vergelijking
 - 3. " " " : onderschikkende voegwoorden die afwijkende hoofdzinsvolgorde regeren
 - 4. " " " : inleidend deel van nevenschikkende voegwoord-groepen
-
- 8.. als 1e cijfer: *interjecties*
 - 0. als 2e cijfer: 'echte' interjecties
 - 1. " " " : substantivisch gebruikte onomatopoeën

- 9.. als 1e cijfer: *restgroepen*
- 9. als 2e cijfer: ('zonder betekenis')
- 0 als 3e cijfer: *anderstalig*
- 1 " " " : niet-lexicale verbindingselementen
- 8 " " " : niet tot steekproef behorende elementen
- 9 " " " : leestekens

HET DERDE CIJFER

Werkwoordsvormen: deelwoorden en infinitieven (2.. als eerste cijfer)

- ..0 infinitief (verbaal)
- ..1 infinitief (substantivisch)
- ..2 onvoltooid deelwoord (onverbogen)
- ..3 onvoltooid deelwoord (verbogen)
- ..4 onvoltooid deelwoord (meervoud)
- ..5 onvoltooid deelwoord (bijwoordelijk)
- ..6 voltooid deelwoord (onverbogen)
- ..7 voltooid deelwoord (verbogen)
- ..8 voltooid deelwoord (meervoud)
- ..9 voltooid deelwoord (bijwoordelijk)

Werkwoordsvormen: persoonsvormen (2.. als eerste cijfer)

- ..1 1e persoon singularis presens
- ..2 2e persoon singularis presens
- ..3 3e persoon singularis presens
- ..4 pluralis presens
- ..5 singularis preteritum
- ..6 pluralis preteritum
- ..7 imperatief zonder pronomen personale
- ..8 imperatief met pronomen personale
- ..9 overige vormen

Andere woordsoorten (eerste cijfer niet 2 en niet 9)

- ..0 basisvorm
- ..1 meervoudsvorm
- ..2 genitief
- ..3 overige verbogen vormen
- ..4 comparatief (onverbogen)
- ..5 comparatief (genitief)
- ..6 comparatief (andere verbogen vormen)
- ..7 superlatief (onverbogen)
- ..8 superlatief (genitief)
- ..9 superlatief (andere verbogen vormen)

4.2.2 Toelichting

0.. substantieven

Via het tweede cijfer zijn vijf onderscheidingen mogelijk.

00.'gewone' substantieven

In het Codeervoorschrift wordt gesteld dat een globaal kenmerk van deze categorie is, dat een lidwoord of een demonstratief of een possessief

als voorbepaling kan optreden. Deze omschrijving blijft vaag, maar toch geeft categorie 00. weinig problemen bij het coderen. Wel moet hier worden opgemerkt, dat niet alle woorden die aan dit globale kenmerk voldoen code 00. krijgen; bijvoorbeeld substantivisch gebruikte adjectieven en infinitieven: *het goede*; *het kopen van kleren*. En omgekeerd, er zijn ook woorden die niet aan dit globale kenmerk voldoen maar toch 00. krijgen, bijvoorbeeld *op grond 00. van*; *van overheidswege 00.* En verder worden ook woorden als *tal*, *paar* en *heleboel* als 'gewoon' substantief gecodeerd: *tal 00. van schilderijen 00.*; *een paar 00. inlichtingen 00.* Het woord *eind* is substantief in *eind 000 januari* terwijl *medio* als adjectief wordt gecodeerd in *medio 100 januari*. Maar deze bijzondere gevallen geven weinig problemen.

01. eigennamen

In het Codeervoorschrift wordt wijselijk geen poging ondernomen deze categorie strikt te definiëren. De kwalificatie 'eigenaam' is vaak subjectief gekleurd en afhankelijk van de voorkennis van de codeerder. Over het algemeen geldt dat alleen woorden die een zeer duidelijk identificerend karakter hebben code 01. krijgen. Bijvoorbeeld - ik noem nu enkele gevallen uit teksten van de overheid - namen van departementen, wetten, nota's, partijen, raden, stichtingen en organisaties. Bij twijfelgevallen geldt als voorwaarde voor codering als eigenaam dat de naam een duidelijk identificerend karakter moet hebben, dus de *ondernemingsraad 00.* tegenover de *Uitkeringsraad 01.* Soort- en merknamen worden met 00. gecodeerd: *een Van Gogh 00.*; *een pakje Rama 00.*

02. adjectivisch gebruikte substantieven

Het gaat hier om de enkele gevallen dat 'gewone' substantieven als voorbepaling optreden: *kunststof 02. voorgevel*; *minimum 02. inkomen*.

08. interjectivisch gebruikte substantieven

Voorbeeld *te droes 08.* (Het woord *te* wordt als voorzetsel gecodeerd.)

09. woorden in zelfnoemfunctie

Niet alleen substantieven maar ook alle andere woorden in zelfnoemfunctie krijgen als code 09. In "Het woord *te*" enkele regels hierboven wordt *te* dus met 09. gecodeerd.

Bij de substantieven zijn via het derde cijfer vier onderscheidingen mogelijk.

..0 basisvorm: *een jaar 000*.

..1 meervoudsvorm: *vijf jaar 001*. Bij de codering 001 wordt dus niet alleen gekeken naar morfologische eigenschappen. Het gaat erom of het woord enkelvoudig of meervoudig wordt gebruikt. Vergelijk ook *onder auspiciën 001 van*.

..2 genitief: *dinsdags 002*; *elk wat wils 002* (maar in *heel wat mans* wordt *mans* als adjectief gecodeerd); *Jan's 012 boek*.

..3 overige verbogen vormen

Het betreft hier substantieven met een verbogen vorm op '-e', op plaatsen waar de datief-functie mogelijk is en waarbij een e-loze vorm bestaat: *bij monde 003 van* maar *op die wijze 000* en *de keuze 000*. Vergelijk *geen sprake 000 van* tegenover *ter sprake 003*. N.B. *van overheidswege* en *in dier voege* krijgen toch 003 ondanks het ontbreken van een e-loze vorm. Maar merkwaardiger wijs moeten wij 000 coderen in *ten behoeve 000 van* en *ten einde 000*.

1.. adjectieven

Via het tweede cijfer zijn vier onderscheidingen mogelijk.

10. 'gewone' adjectieven

Hiertoe worden gerekend zowel de adjectieven die attributief gebruikt worden, *een extra 10. uitkering*, als de adjectieven die predikatief gebruikt worden, dat wil zeggen als kern van een naamwoordelijk deel van het gezegde of als bepaling van gesteldheid: *hij is heel eerlijk 10.*; *ik vind hem heel aardig 10*. De omschrijving 'predikatief gebruikt' is heel ruim. Vergelijk: *ik heb het nodig 10.*; *dit staat mooi 10.*; *hij schopt het ver 10.*; *dat zit wel snor 10*. En verder wordt in uitdrukkingen als *een zo groot mogelijke overwinning*, groot als gewoon adjectief beschouwd, maar *mogelijke* als adverbiaal gebruikt adjectief. Een 'gewoon' adjectief vinden we in: *hij is voornemens 10*; *zich bezig 10. houden met*; *de grondstoffen raken uitgeput 10.*; *wij zijn hiermee gebaat 10*.

12. substantivisch gebruikte adjectieven

Het gaat hier in het algemeen om verbogen adjectieven die substantivisch zijn gebruikt: *de goeden 12.*; *een jongere 12*. Dus in het alge-

meen houdt code 000. Wel krijgen de niet verbogen substantivisch gebruikte namen van talen en kleuren code 12.: *dit lijkt op blauw 12.*; *hij spreekt Engels 12.* En ook is 12. van toepassing op een feest voor jong 12. en oud 12.

15. adverbiaal gebruikte adjectieven

Dit zijn, negatief gedefinieerd, adjectieven die niet attributief of predikatief of zelfstandig worden gebruikt. Deze categorie onderscheidt zich van de bijwoorden, in die zin dat het hier gaat om woorden die ook als attributief gebruikt adjectief kunnen optreden. Vergelijk: *hij heeft de wedstrijd natuurlijk 15. gewonnen* en *hij heeft de wedstrijd uiteraard (bijwoord) gewonnen*. Overigens worden graadaanduidende bijwoorden altijd als bijwoord gecodeerd. In ruim 1 miljoen daarentegen krijgt ruim de code 15. Nog enkele voorbeelden over het onderscheid 10. en 15.: *de enig 15. juiste 10. oplossing*; *een zo spoedig 10. mogelijke 15. terugkeer*; *zo spoedig 15. mogelijk 15. bericht ik u*. En in *te weinig rekening houden met zijn bij weinig* beide mogelijkheden te verdedigen.

18. interjectivisch gebruikte adjectieven

goed 18., dat is duidelijk.

Via het derde cijfer zijn bij de adjectieven tien onderscheidingen mogelijk.

..0 basisvorm

..1 meervoudsvorm

Dit komt uiteraard alleen voor bij substantivisch gebruikte adjectieven: *de goeden 121*. Voorwaarde hierbij is dat het woord eindigt op '-en', en niet in de vergrotende of overtreffende trap staat. We krijgen 123 in *die lelijke moeten meteen weg*.

..2 genitief

iets moois 102; zaliger 102 nagedachtenis.

..3 overige verbogen vormen

Dit zijn voornamelijk de adjectieven met buigingsuitgang '-e': *rode 103 rozen*. (Maar *rose 100 bloemen* omdat hier geen sprake is van een basisvorm); *een afvallige 123*; *een zo goed mogelijke 153 uiteenzetting*. En verder vallen hier ook onder oude datievormen als: *in goeden 103 doen*; *te goeder 103 trouw*.

De volgende onderscheidingen komen vrijwel alleen voor bij adjectieven.

..4 comparatief (onverbogen)

hij is nauwkeuriger 104; hij legt nauwkeuriger 154 uit. N.B. Ook eerder en nader krijgen een 4 als derde cijfer, hoewel de basisvorm ontbreekt.

..5 comparatief (genitief)

iets mooiers 105.

..6 comparatief (andere verbogen vormen)

een betere 106 opzet; de ouderen 126 onder ons

Via de cijfers ..7 en ..8 en ..9 worden dezelfde onderscheidingen gemaakt bij de superlatieven: *de grootst 107 mogelijke overwinning; op het heetst 127 van de dag; eerst 157 A en vervolgens B (vervolgens is een bijwoord).*

Woorden van het type *alleruiterst* krijgen een ..0, evenals *dichtstbijzijnd*.

Minstens en *hoogstens* zijn bijwoorden. Als voorbeelden voor ..9

gelden (..8 bij adjectieven komt in de gecodeerde teksten niet voor):

de laatste 109 man; de laatsten 129; ten nauwste 159.

2.. werkwoordsvormen

Via het tweede cijfer wordt hier onderscheid gemaakt tussen infinitieven en deelwoorden enerzijds (tweede cijfer 0, 1, 2 of 3) en persoonsvormen anderzijds (tweede cijfer 4, 5, 6, 7 of 8). Er worden vier onderscheidingen gemaakt die voor alle werkwoordsvormen gelden, dus zowel voor infinitieven en deelwoorden als voor persoonsvormen.

20. 24. intransitief gebruikte werkwoorden

Deze categorie is negatief gedefinieerd: een werkwoordsvorm is intransitief als hij niet transitief of reflexief is gebruikt, en als het geen hulp- of koppelwerkwoord is.

21. 25. transitief gebruikte werkwoorden

Dit zijn de werkwoordsvormen waar een direct object bij staat. Hierbij moet worden vermeld, dat het begrip 'direct object' ruimer is gedefinieerd dan het traditionele lijdend voorwerp. Ook in de volgende zinnen wordt het werkwoord als transitief gecodeerd: *de stad uitkomen 21.; zijn gang gaan 21.; ik heb 25. het koud; dat kost 25. twee gulden; ik zag 25. hem denken.*

Van transitief gebruik is geen sprake als het complement van het werkwoord een bijwoordelijke bepaling van tijd is, *de kabinetsformatie duurde 24. negen maanden*, of als de bijzin een voorzetselvoorwerp is: *hij rekende 24. erop dat hij minister zou worden.*

Wel kunnen werkwoordsvormen die geen direct object bij zich hebben, als transitief gecodeerd worden: *de maatregelen worden gewijzigd 21.; het*

land bleek verlaten 21.; een verveelde 21. blik; het inpakken 21. van dozen. Tot slot zij nog vermeld dat in de zin *dit codeersysteem is gemakkelijk toe te passen, zo meende 24. hij*, het gebruik van zo verhindert dat *meende* als transitief wordt gecodeerd.

22. 26. reflexief gebruikte werkwoorden

Hier moet worden opgemerkt, dat bij de keuze tussen transitief of reflexief altijd de voorkeur wordt gegeven aan transitief. Dus een werkwoordsvorm met een reflexivum en een direct object wordt transitief gecodeerd: *de vraag laat zich stellen 21.; hij herinnerde 25. zich dat niet.*

23. 27. hulp- en koppelwerkwoorden

Tot de hulpwerkwoorden worden ook gerekend de hulpwerkwoorden van modaliteit. Maar ook andere werkwoorden kunnen als hulpwerkwoord optreden: *hij probeert 27. de strijd te beëindigen; het beloofde 27. een mooie dag te worden.* Wel moet bij het coderen als hulpwerkwoord altijd een afhankelijke infinitief aanwezig zijn.

Over de koppelwerkwoorden nog het volgende. Naast de traditionele koppelwerkwoorden worden ook de werkwoorden in de volgende zinnen als zodanig gecodeerd: *de naam luidt 27. als volgt; de zaal raakt 27. vol.* Merkwaardigerwijs is er in *de zaal loopt 24. vol* volgens het Codeervoor-schrift geen sprake van een koppelwerkwoord.

Als laatste tweede cijfer voor de persoonsvormen moet nog genoemd worden:

28. interjectivisch gebruikte werkwoorden

Naast de vier genoemde onderscheidingen voor het tweede cijfer na 2.. wordt code 28. gebruikt voor enkele imperatiefvormen die semantisch niet meer als echte imperatieven zijn te beschouwen: *dat heb je mooi gedaan, hoor 287.*

De betekenis van het derde cijfer bij werkwoordsvormen is afhankelijk van het onderscheid tussen infinitieven en deelwoorden (20., 21. 22., 23.) en persoonsvormen (24., 25., 26., 27., 28.).

Voor de infinitieven en deelwoorden zijn er tien mogelijkheden.

..0 infinitief (verbaal)

met studeren 200 bereik je weinig; hij was een boek aan het lezen 210.

..1 infinitief (substantivisch)

al dat studeren 201 haalt niets uit; mijns inziens 201.

..2 onvoltooid deelwoord (onverbogen)

zich bevindend 222.

..3 onvoltooid deelwoord (verbogen)

iets boeiends 203; zijnde 233.

..4 onvoltooid deelwoord (meervoud)

de studerenden 204.

..5 onvoltooid deelwoord (bijwoordelijk)

hij is voortdurend 205 bezig; het geheel overziende 215, sprak hij ...

N.B. Bij een mogelijke codering met ..2 of ..3 moet dus eerst bepaald worden of het woord niet bijwoordelijk is gebruikt.

..6 voltooid deelwoord (onverbogen)

een verouderd 206 pand; het schilderij is weggestopt 216 in een la; gelet 206 op ...; gezien 216 ...; zij hebben zich verheugd 216 getoond 226; hij is ziek geweest 236.

..7 voltooid deelwoord (verbogen)

het afgekeurde 217 verslag; de ondergetekende 207.

..8 voltooid deelwoord (meervoud)

de ondergetekenden 208.

..9 voltooid deelwoord (bijwoordelijk)

bepaald 209 niet; geboeid 219 keek hij toe; hij reageerde geschrokken 209.

N.B. Bij een mogelijke codering met ..6 moet dus eerst bepaald worden of het woord niet bijwoordelijk is gebruikt. De voltooiden deelwoorden in constructies als *over het algemeen gesproken 206* en *op de keper beschouwd 206* worden niet als bijwoordelijk gecodeerd, omdat de deelwoordconstructie in het geheel als bijwoordelijke bepaling optreedt. Vergelijk ook *afgezien 206 van*.

Bij het coderen als deelwoord doen zich nogal eens twijfelgevallen voor omdat de deelwoorden, en vooral de onvoltooiden, soms moeilijk te onderscheiden zijn van adjectieven of bijwoorden. Volgens het Codeervoorschrift moet een vorm als deelwoord worden gecodeerd, als dat op grond van morfologische eigenschappen ook inderdaad mogelijk is. Dus *het beperkte 217 bereik*. Van dit voorschrift wordt afgeweken als het woord in kwestie semantisch niets meer te maken heeft met het werkwoord waarvan het is afgeleid, of als het bijbehorende werkwoord in onbruik is geraakt. Dus *het volgende 203* tegenover *het voorgaande 123* en *het voorafgaande 203* tegenover *enkele vooraanstaande 103 politici* (in de betekenis van 'belangrijke'), en *dit is geoorloofd. 100*.

Bij codering als deelwoord geldt altijd dat er een bijpassende persoonsvorm of infinitief moet zijn, dus *geschreven 216* en *ongeschreven 100 richtlijnen* en *werkenden 204* tegenover *werkzoekenden 121*. Merkwaar-

digerwijs geldt dit niet voor *bepaald 209 niet*.

Verder kunnen er nog problemen optreden bij woorden die formeel identiek zijn aan deelwoorden, maar in een andere functie optreden: *aangezien* kan voegwoord zijn; *aangaande* en *gezien* kunnen optreden als voorzetsel (*betreffende* moet altijd als voorzetsel worden gecodeerd).

Een apart probleem vormen woorden als *bepaald*, *verschillend* en *voldoende*.

Deze woorden kunnen ook als indefinitum gecodeerd worden, vergelijk *een bepaald geval* (indefinitum) tegenover *bepaald 209 niet!* (bijwoordelijk gebruikt deelwoord) of *het bepaalde 217* (verbogen voltooid deelwoord).

Soms kan er ook twijfel bestaan over het coderen als substantivisch gebruikt infinitief en 'gewoon' substantief, vergelijk: *het varen 201* en *het streven 000*.

Voor het derde cijfer bij persoonsvormen (24., 25., 26., 27., 28.) zijn er negen mogelijkheden.

..1 1e persoon singularis presens

(ik) dank 241 je.

..2 2e persoon singularis presens

bent 272 u daar geweest?

..3 3e persoon singularis presens

(het) gaat 243 wel.

..4 pluralis presens

wij kunnen 274 niets doen.

..5 singularis preteritum

hij zou 275.

..6 pluralis preteritum

wij gingen 246.

..7 imperatief zonder pronomen personale

De naam is hier verwarrend. Het gaat blijkens het Codeervoorschrift om subjectloze imperatieven. Dus *pak 257 ze* valt wel onder deze categorie. Ook in de volgende categorie wordt met 'pronomen personale' het subject bedoeld.

..8 imperatief met pronomen personale

blijven 248 jullie maar hier.

..9 overige vormen

Hiermee worden bedoeld de conjunctieven en andere afwijkende vormen: *het ga 249 je goed; het zij 279 zo; het ware 279 te wensen; zegge 259 en schrijfve 259.*

3.. pronomina A

Hier zijn via het tweede cijfer vier hoofdonterscheidingen mogelijk: personalia, possessiva, reflexiva en demonstrativa. De laatste drie soorten pronomina kunnen worden onderscheiden naar zelfstandig en bijvoeglijk gebruik. Bij dit onderscheid zijn steeds de even getallen genomen voor zelfstandig gebruik en de oneven voor bijvoeglijk. Dit geldt ook voor de pronomina B (4..).

30. personalia

Deze categorie behoeft weinig toelichting. Wel moet worden opgemerkt, dat *je* in de betekenis van *men* als indefinitum (zie 44.) wordt gecodeerd. Moeilijkheden kan het woord *het* geven. Dit woord krijgt alleen de codering 30. als het betrekking heeft op een onzijdig woord of als het vervangen kan worden door *hij* of *zij*: *het 30. zijn schoften*.

Via het derde cijfer maken we onderscheid tussen: basisvorm (300), genitief, *gedenk mijner 302*, en andere verbogen vormen (303). De codering 303 is gereserveerd voor personalia die niet in subjectsfunctie zijn gebruikt: *heb jij 300 hen 303 gezien?* Het gaat hier om de functie en niet om de vorm, dus ook: *heb jij ze 303 gezien?*

32. 33. possessiva

Bij zelfstandig gebruik volgt de codering 32.; bij bijvoeglijk gebruik de codering 33. Het derde cijfer onderscheidt hier: basisvorm ..0, meervoud ..1, genitief ..2 en overige vormen ..3. Voorbeelden: *de onzen 321; al het uwe 323; mijns 332 inziens; dit zijn jouw 330 schoenen niet, maar de mijne 333*. N.B. Uit het laatste voorbeeld blijkt dat schijnbaar zelfstandig gebruik toch als bijvoeglijk wordt gecodeerd.

34. 35. reflexiva

Hier onderscheidt het derde cijfer: basisvorm ..0, genitief ..2 en overige verbogen vormen ..3.

Ook het woord *zelf* krijgt deze codering en wordt als zelfstandig opgevat, tenzij het direct volgt op het woord waar het bijhoort. Vergelijk: *hij zei het zelf 340; hij zelf 350 zei het. Hijzelf*, aaneengescreven, krijgt 340, want ook samenstellingen met *zelf(de)* worden opgevat als reflexivum. Ook *eigen* krijgt 34. of 35.; het blijft uiteraard adjectief in: *zich iets eigen 100 maken*. Nog enkele voorbeelden: *vergis je 340 niet; elkaars 342 mening respecteren* (*elkaar* moet hier als zelfstandig gecodeerd worden, want alleen de genitief-s is hier verantwoordelijk voor het bijvoeglijke element; vergelijk de opmerking over *diens mening* onder 36. en 37.); *eenzelfde 350 opvatting; de commissie zelve 353*.

36. 37. demonstrativa

Ook hier onderscheidt het derde cijfer: basisvorm ..0, meervoudsvorm ..1, genitief ..2 en overige verbogen vormen ..3.

Naast woorden als *dergelijk*, *zodanig*, *zoiets* en *hij* in de betekenis 'diegene' wordt ook voor de lidwoorden *de* en *het* deze codering gebruikt. Voorbeelden: *zulks* 360; *degenen* 361 *die*; *des* 362 *te beter*; *diens* 362 *mening* (bij genitiefvormen wordt gekeken naar de grondvorm als het gaat om zelfstandig of bijvoeglijk gebruik); *en dergelijke* 363 (ook *degene* krijgt 363, hoewel *degeen* niet voorkomt); *het* 370 *boek*; *de heer des* 372 *huizen*; *dergelijke* 373 *opvattingen*. Schijnbaar zelfstandig gebruik wordt hier - in tegenstelling tot het voorschrift onder possessiva - toch als zelfstandig gecodeerd: *Breng je dit 370 boek mee of dat 360?* Nog twee voorbeelden over het onderscheid zelfstandig-bijvoeglijk: *de resultaten zijn zodanig 360 dat ...; iets dergelijks 372.*

4.. pronomina B

Hier zijn via het tweede cijfer vijf hoofdonderscheidingen mogelijk: interrogativa, relativa, indefinita, cardinalia en ordinalia. Binnen elke categorie wordt weer het onderscheid zelfstandig-bijvoeglijk gehanteerd. Net als in 3.. zijn de even tweede cijfers gereserveerd voor het zelfstandig gebruik en de oneven tweede cijfers voor het bijvoeglijk gebruik.

40. 41. interrogativa

Via het derde cijfer worden onderscheiden: basisvorm ..0, genitief ..2 (voor bijvoeglijk gebruikte interrogativa komt dit zelden voor) en overige verbogen vormen ..3. Voorbeelden: *wie* 400 *komen er?*; *door wiens* 402 *toedoen?*; *welke* 403 *heb jij?*; *welke* 413 *boeken?* Een interrogativum komt ook voor in een afhankelijke vraag: *ik weet niet wie 400 er komen*. Het woord *wie* in de constructie *wie ... ook* krijgt eveneens deze codering. Verder krijgt ook het woord *wat* in een uitroep 40. of 41.: *wat 400 ben ik blij!*; *wat 410 een mensen!*

42. 43. relativa

Via het derde cijfer ook hier: basisvorm ..0, genitief ..2 en overige verbogen vormen ..3. Voorbeelden: *het antwoord dat 420 je gaf* (in "het antwoord dat alles in orde zou komen" is *dat* voegwoord); *de man wiens* 422 *hoed in het water viel*; *de maatregelen welke 423 zijn genomen*. Bijvoeglijk gebruik komt zelden voor, alleen bij *welke*: *ijshockey, van welke 433 sport wij niets begrijpen*.

44. 45. indefinita

Deze categorie is helaas zeer heterogeen. In het Codeervoorschrift worden vier soorten indefinita onderscheiden. Enkele voorbeelden:

- a. categoriale indefinita: alles, iedere, al (haar vrienden);
- b. existentiële indefinita: iemand, iets, elk wat (wils), verscheidene;
(Tot deze groep behoort ook 'een' als onbepaald lidwoord.)
- c. negatieve indefinita: geen, niets;
- d. kwantitatieve indefinita: veel, min (of) meer, genoeg, onvoldoende, beide, generlei, ander.

Via het derde cijfer kan hier, naast de onderscheidingen genoemd bij de andere pronomina, ook nog worden aangegeven of het woord in de comparatief of superlatief staat, en wel in onverbogen (..4 en ..7) of in verbogen vorm (..6 en ..9). Voorbeelden: *een 440 van de aardigste jongens die ik ken*; *tallozen 441*; *ieders 442 opvatting*; *dat is voldoende 443* (ook *sommige* wordt als verbogen gecodeerd, hoewel de basisvorm hier ontbreekt); *hij komt niet meer 440 dan vroeger* (in "hij komt niet meer" is *meer* een bijwoord); *hij is altijd de mindere 446*; *de meest 447 boeiende tijdschriften* maar in "de meest ongelooflijke verhalen" is *meest* bijwoord, omdat dit woord hier een adjectief bepaalt); *ten minste 449*; *allerhande 450 koopwaar*; *de lotgevallen ener 452 dame*; *enkele 453 opmerkingen*; *de zoveelste 453 keer*; *meer 454 dan zeven dagen*; *meerdere 456 verschijnselen*; *het meest 457 aantrekkelijke voorstel*; *de meeste 459 mensen*.

Soms is de grens tussen indefinita en adjectieven moeilijk te trekken. In *Woordfrequenties* staan bijvoorbeeld *aanzienlijke* en *overige* steeds als adjectief gecodeerd. En bij *diverse* zijn de codes 103 en 453 mogelijk.

46. 47. cardinalia

Via het derde cijfer worden hier onderscheiden: basisvorm ..0, meervoudsvorm ..1 en overige verbogen vormen ..3. Voorbeelden: *4,5 460*; *honderden 461*; *tegen vieren 463*; *artikel 7 470*; *duizenden 471 mensen*; *anderhalve 473 man*.

Cijfers en getallen in woorden waar spaties in voorkomen, worden, als één telwoord beschouwd. Een telwoord wordt bijvoeglijk gecodeerd als het afhankelijk is van een substantief. Dus *vijftien 470 uur dertig 460*. 'Bijvoeglijk gebruik' geldt ook voor het telwoord dat afhankelijk is van het bijwoord *er*: *we zagen er een stuk of twaalf 470* naast *we zagen een stuk of twaalf 460*.

N.B. Als substantief worden gecodeerd: *hij kreeg een vijf 000*; *met veel vijven 001 en zessen 001*; *een dertiger 000*. De woorden *getweën* en *eenmaal* zijn bijwoorden.

48. 49. ordinalia

Na de uitleg onder cardinalia kan hier worden volstaan met enkele voorbeelden: *hij was de derde* 480; *zij kwamen als eersten* 481; *hij kwam als eerste* 483. Het woord *eerste* wordt opgevat als verbogen vorm naast de basisvorm 'eerst'; vandaar *de eerste* 493 *plaats* tegenover *de derde* 490 *plaats*. N.B. *laatste* is geen ordinale: *de laatste* 109 *plaats*.

Het zal duidelijk zijn dat bij een aantal woorden die thuishoren onder pronomina A en B, verschillende coderingen mogelijk zijn. Hier enkele voorbeelden:

die 360	die daar
370	die man
420	de man die

een 440	het een of het ander
450	een boek (ene 453 Jansen)
460	geef maar één
470	wij hebben er één

(Ook de codering 100 kan voorkomen bij één: *zich één* 100 *voelen met*.)

het 300	(over een kind) het huilt
303	geef het te drinken
370	het kind
440	(het huis brandt) heb je het gezien?

je 300	je doet maar
303	ik geef het je
330	je boek
340	overwerk je niet
440	je kunt nooit weten

wat 400	wat voor boeken zijn dat?
410	wat een boeken!
420	wat hij doet, weet ik niet
440	dat is heel wat
450	zij is wat zwijgzaam

wie 400	wie er komt, weet ik niet
420	wie er komt, zal er geen spijt van hebben

5.. bijwoorden

Via het tweede cijfer onderscheiden wij acht categorieën: 50. 'gewone' bijwoorden; 51. aanwijzende en onbepaalde bijwoorden; 52. vragende bijwoorden; 53. betrekkelijke bijwoorden; 54. voornaamwoordelijke bijwoorden; 55. vragende voornaamwoordelijke bijwoorden; 56. betrekkelijke voornaamwoordelijke bijwoorden en 58. interjectivisch gebruikte bijwoorden.

50. 'gewone' bijwoorden

Deze klasse wordt negatief gedefinieerd: woorden die niet tot een van de andere categorieën onder 5.. behoren, krijgen code 50.

Enkele voorbeelden: *heel 500 groot; zeer 500 belangrijk; de kachel is uit 500; dat is even 500 mooi* (even als tegenovergestelde van 'on-even' is een bijvoeglijk naamwoord); *dicht 500 bij huis; midden 500 in het bos; twee jaar geleden 500; heen 500 en weer 500; t.z.t. 500; terwille 500 van* (maar *ter 603 wille 003 van*). Tot slot nog enkele voorbeelden van homoniemen:

links links 500 achter het huis - hij is links 100 (linkshandig, of 'in de politiek')

eens eens 500 te 500 meer 500 - het eens 100 zijn met

eerst eerst 500 nu besef ik ... - hij kwam het eerst 157

alleen ik zei alleen 500 maar ... - ik ga alleen 100

thuis ik ben niet thuis 500 - dat heeft ze van thuis 000

Via het derde cijfer onderscheiden wij alleen bij de 'gewone' bijwoorden: basisvorm ..0, overige verbogen vormen ..3 (de verbogen vormen bij de andere bijwoorden krijgen merkwaardigerwijs geen ..3: *dusverre 510; zoverre 510*), comparatief ..4, superlatief ..7 en superlatief in andere verbogen vormen dan de genitief ..9. Enkele voorbeelden: *ten enenmale 503; vaker 504; het dichtst 507 bij; ten zeerste 509*.

510 aanwijzende en onbepaalde bijwoorden

Voorbeelden: *er, dan, nooit, dus, zodoende, daarentegen* enz.

N.B. De woorden *eens, weleens, steeds, aldoor* krijgen code 500. Ook 510 krijgen de bijwoorden met een historische genitief of datief: *derhalve, indertijd, uiteraard, mitsdien*.

520 vragende bijwoorden

Tot deze categorie behoren de woorden: *wanneer, waar, waarom* (om welke reden) *hoe* en bijwoorden met *hoe* samengesteld, *in hoeverre 520*.

N.B. *hoeveel* krijgt altijd codering 440 of 450 (indefinitum).

530 betrekkelijke bijwoorden

Hier gaat het om bijwoorden die een bijzin inleiden: *de stad waar*

530 ik woon; de reden waarom 530 ik kom. Ook *dat* kan deze codering krijgen: *in het jaar dat 530 de oorlog uitbrak*.

540 voornaamwoordelijke bijwoorden

Bij de voornaamwoordelijke bijwoorden worden dezelfde onderscheidingen gemaakt als bij de bijwoorden. Bij 540 gaat het om de aanwijzende of onbepaalde voornaamwoordelijke bijwoorden. De vragende en betrekkelijke voornaamwoordelijke bijwoorden worden gecodeerd met respectievelijk 550 en 560. Als het voornaamwoordelijk deel gesplitst voorkomt, dan krijgt het eerste deel toch de code 540 (550 of 560). Het tweede deel krijgt dan de code 610. Enkele voorbeelden van (aanwijzende of onbe-

paalde) voornaamwoordelijke bijwoorden: *ermee, daarvan, hierover, daartussenin.*

550 vragende voornaamwoordelijke bijwoorden
waarheen?; waarbij?

560 betrekkelijke voornaamwoordelijke bijwoorden
de plaats, waarbij 560 het gebeurde; het doel waarvoor 560 je werkt.

580 interjectivisch gebruikte bijwoorden
Enkele voorbeelden: *ik doe maar 580 wat; nou 580 en of; zo 580, die is ver heen.*

Ook in categorie 5.. komen een aantal woorden voor waarbij zich verschillende coderingsmogelijkheden voordoen. Enkele voorbeelden:

meer	444	hij komt niet meer dan vroeger
	454	hij heeft meer boeken dan ik
	500	hij komt niet meer

(meer wordt uiteraard als substantief gecodeerd in: *het meer 000.*)

waar	520	waar ben je?
	530	de plaats waar ik woon
	550	waar ga je heen? (<i>heen</i> krijgt dan 610)
	560	een plaats waar je heen kunt (<i>heen</i> krijgt dan 610)
	710	als onderschikkend voegwoord in de betekenis 'omdat' of 'hoewel'

(*waar* kan ook nog gecodeerd worden als substantief; als 'gewoon' of ad-verbiaal of interjectivisch gebruikt adjectief en als persoonsvorm.)

waarom	520	waarom doe je dat?
	530	de reden waarom
	550	waarom wikkel je dat papier?
	560	de vraag waarom ik het doe, zal ik straks beantwoorden

(*waarom* wordt als substantief gecodeerd in: *het waarom 000.*)

6.. pre- en postposities

Hier worden via het tweede cijfer zeven categorieën onderscheiden.

Allereerst de traditionele voorzetsels, 60. (Alleen hier speelt het derde cijfer nog een rol: basisvorm ..0, verbogen vorm ..3.)

Verder wordt deze code gebruikt voor vier categorieën woorden, die in de woordbenoeming soms ook tot de bijwoorden worden gerekend:

610 tweede deel van gesplitste naamwoordelijke bijwoorden;

620 niet-verbaal deel van samengestelde werkwoorden;

630 tweede deel van gescheiden voorzetselgroepen;

640 achterzetsels.

Daarnaast zitten in 6.. nog twee categorieën voorzetsels die syntactisch gezien merkwaardig gebruikt worden:

650 te bij een infinitief en een voorzetsel met een te+infinitief-constructie als rest;

660 voorzetsels met een door onderschikkend voegwoord ingeluide bijzin als rest.

Hieronder van elke categorie enkele voorbeelden.

60. voorzetsels

bij 600 monde van 600; ter 603 aangehaalde plaatse; tot 600 aan 600 de deur (vergelijk voor 500 in 600 de straat). Voor het onderscheid tussen voorzetsel en bijwoord geldt het volgende criterium: een voorzetsel moet altijd een in de zin aan te wijzen 'rest' hebben, vandaar dat *met* als bijwoord gecodeerd wordt in *patat met 500*.

Andere voorbeelden van woorden die men niet direct tot deze categorie zou rekenen: *gezien 600 het feit dat betreffende 600 het voorstel overeenkomstig 600 de voorschriften is gehandeld; exclusief 600 bedieningsgeld; halverwege 600 de derde etage.*

610 tweede deel van gesplitste voornaamwoordelijke bijwoorden

Daar 540 doe ik het mee 610; waar 550 kom jij vandaan 610?

Het woord *in* krijgt ook 610 in: *bier om in te bijten* omdat deze zinswending, aldus het Codeervoorschrift, betekent: "bier waarin je kunt bijten".

620 niet-verbaal deel van samengestelde werkwoorden

Deze codering wordt alleen toegepast als het voorzetsel niet aan het werkwoord vast zit: *hij belde 255 mij op 620; ik doe 241 daar niet aan mee 620.*

Behalve een enkele uitzondering (*het duurt 243 voort 620; hij reist de wereld rond 620*) kan een woord alleen 620 krijgen als het in ander verband met 610 kan worden gecodeerd. Dus geen 620 in: *hij gaat dood 100; hij komt terug 500.*

630 tweede deel van gescheiden voorzetselgroepen

tegen 600 de voorschriften in 630; durf jij erdoor 540 heen 630? (vergelijk: durf je er 540 door 610 heen 630?); tot 600 aan 600 de deur toe 630; tot 600 de deur aan 630 toe 630.

640 achterzetsels

mijzelf inbegrepen 640; ze durfden het huis niet in 640.

650 te bij een infinitief en een voorzetsel met een te+infinitief-constructie als rest

door 650 hard te 650 lopen haalde hij de trein; teneinde 650 hieraan uitvoering te 650 kunnen geven.

660 voorzetsels met een door onderschikkend voegwoord ingeluide bijzin als rest

tegen 660 dat we weer thuis zijn; voor 660 als je je weer wat beter voelt; in plaats van 660 dat hij komt. Vergelijk in plaats van 650 te komen.

Ook in categorie 6.. komen woorden voor waarbij verschillende coderingen mogelijk zijn. Enkele voorbeelden:

rond	000	ze huppelen in het rond
	100	de bal is rond
	150	tien seconden rond
	500	rond duizend gulden
	600	rond het huis
	620	hij hing maar wat rond
	640	hij liep het huis rond
te	500	te gek
	600	te Amsterdam
	650	door te huilen bereik je niets
toe	500	de deur is toe
	610	dat doet er niets toe
	620	hij gaat er naar toe
	630	tot daar aan toe
uit	253	hij uit bezwaren
	263	hij uit zich moeilijk
	500	de kachel is uit
	600	uit de wind
	610	waar concludeer je dat uit?
	620	kijk uit!
	630	hij kneep ertussen uit
	640	mijn huis uit!

7.. conjuncties

Via het tweede cijfer worden onderscheiden: nevenschikkende voegwoorden, 700; onderschikkende voegwoorden, 710; voegwoorden van vergelijking, 720; onderschikkende voegwoorden die afwijkende hoofdzinsvolgorde regeren, 730 en inleidend deel van nevenschikkende voegwoordgroepen, 740. Er zijn geen onderscheidingen via het derde cijfer. Hieronder van elke categorie enkele voorbeelden.

700 nevenschikkende voegwoorden

thee of 700 koffie?; al dan 700 niet. Ook of in balansschikking krijgt 700: er ging geen dag voorbij of 700 ...

710 onderschikkende voegwoorden

zolang 710 er leven is, is er hoop (maar de kruik gaat zolang 510 te water tot hij breekt); hij is niet zo slim als 710 jij denkt; voorzover 710 ik weet, komt hij; nu 710 ik beter ben, kan ik komen.

720 voegwoorden van vergelijking

niets dan 720 ellende; hij is niet zo slim als 720 jij; als 720 vakman is hij zeer geslaagd. Maar code 720 wordt ook gebruikt in: twee plus 720 twee; alle leden alsmede 720 de voorzitter; behalve 720 de voorzitter, was iedereen aanwezig; hoewel 720 niet rancuneus, moet ik toch opmerken ...

Het criterium voor 720 is: het niet voorkomen van een bijzin als rest.

Vergelijk: *zo 720 ja en zo 710 u ja zegt, dan ...* Zie ook: *dat kan indien 720 nodig als 710 volgt worden verklaard.*

730 onderschikkende voegwoorden die afwijkende hoofdzinsvolgorde regeren.

Voorbeeld: *al 730 mag ik haar niet, ik zal haar toch begroeten.* (Vergelijk de volgorde van *ik* en *zal* in: *ondanks 710 het feit dat ik haar niet mag, zal ik haar toch begroeten.*) Ook *als* krijgt 730 wanneer de persoonsvorm van de afhankelijke bijzin (!) er onmiddellijk op volgt: *hij keek als 730 was er niets aan de hand.* De benaming 'afwijkende hoofdzinsvolgorde' is dus niet helemaal correct.

740 inleidend deel van nevenschikkende voegwoordgroepen

noch 740 zij noch 700 wij; zowel 740 ... als 700. Het tweede, en eventueel het derde deel, krijgt altijd de codering 700. Uiteraard moet er wel sprake zijn van voegwoorden. Bijwoorden vinden we in: *enerzijds 500 ... anderzijds 500; niet 500 alleen 500 ... maar 700 ook 500.*

Ook in categorie 7.. komen woorden voor die verschillende codeermogelijkheden hebben. Enkele voorbeelden:

al	440	al met al
	450	al die jaren
	500	al te gek; al dan niet
	730	al mag ik haar niet, ik zal haar toch begroeten
als	700	zowel Jan als Piet
	710	Jan is niet zo sterk als ik dacht
	720	Jan is even sterk als Piet
	730	hij zong als had hij nooit anders gedaan
daar	510	hij is daar
	540	daar kan ik het niet mee doen
	710	daar de minister niet geantwoord heeft, herhaal ik de vraag
dan	510	dan doe ik het
	700	of hij al dan niet gelijk heeft ...
	710	hij is groter dan ik dacht
	720	anders dan anders
dat	360	dat moet je niet doen
	370	dat boek
	420	het huis dat daar staat
	530	het jaar dat de oorlog uitbrak
	710	hij zei dat het goed was

of	700	een stuk of drie; koffie of thee; nou en of
	710	weet jij of hij komt?
	740	<i>of a of b</i> (het tweede <i>of</i> krijgt 700)
sinds	600	sinds 1900
	660	sinds dat het zo duur is geworden, koop ik het niet meer
	710	sinds het zo duur geworden is, koop ik het niet meer
tot	500	een schurk tot en met
	600	tot hier en niet verder
	710	we zullen wachten tot hij komt
zo	510	dat is zo
	580	zo, dat was dat
	710	zo hij mocht komen, dan zal ik je dat laten weten
	720	zo ja, dan kunnen we doorgaan

8.. interjecties

Interjecties worden in het Codeervoorschrift omschreven als woorden die buiten elk zinsverband vallen. Hier zijn twee onderscheidingen mogelijk.

800 'echte' interjecties

integendeel 800; ja 800. Ook het antwoord *ja* krijgt 800, maar dit woord krijgt 500 in *zo ja 500*. Want voor code 800 is het noodzakelijk, dat het woord buiten het zinsverband valt. Meestal is dit al aangegeven door de interpunctie. Vergelijk *ga nou, alsjeblieft 800!* en *wil je alsjeblieft 500 komen?*

810 substantivisch gebruikte onomatopoeën

De naam van de categorie geeft voldoende uitleg: *een zacht grrrr 810 deed hem opkijken.*

9.. restgroepen

Hier zijn vier onderscheidingen mogelijk.

990 anderstalig

Het gaat hier om niet-Nederlandstalige uitdrukkingen en citaten: *hij sprak van een 'gentlemen's agreement' 990.* (De uitdrukking tussen aanhalingstekens wordt hier als één woord gecodeerd. De spatie voor 'agreement' is in dit codeersysteem een 'oneigenlijke spatie'; zie pagina 98.)

991 niet-lexicale verbindingselementen

Deze codering wordt alleen voor gesproken taal gebruikt, en wel voor de verbindende 'n' tussen klinkers (*toen hoorde n 991 ie een klap*), voor onverstaanbare tekstgedeelten en voor aarzelingen (*eh*).

Tot slot nog twee coderingen die te maken hebben met het codeerproces zelf.

998 niet tot de steekproef behorende elementen

Deze codering wordt gebruikt voor woorden die in de steekproef zijn meegenomen, maar er eigenlijk buiten vallen (bijvoorbeeld datering en aanduiding van plaats van herkomst).

999 leestekens

Deze codering geldt voor de leestekens die - aldus het Codeervoorschrift - ten onrechte door de computer als woord zijn beschouwd, bijvoorbeeld het uitroepteken of '...'.

Het vierde cijfer

Hoewel het C_3C -systeem in *Woordfrequenties* als drie-cijfercodering is gepresenteerd, wordt toch in twee gevallen een codering van vier cijfers toegepast, namelijk bij voornaamwoordelijke bijwoorden wanneer de delen los van elkaar staan, en bij samengestelde werkwoorden wanneer het niet-werkwoordelijk deel los van het werkwoord staat. Bijvoorbeeld: *hij voegde daar nog iets aan toe*. In zo'n geval worden de bij elkaar behorende woorden gecodeerd met het vierde cijfer 0 en 5 (en eventueel 1 en 6, 2 en 7, wanneer dit verschijnsel vaker voorkomt in een zin). Dus: *hij voegde 2550 daar 5401 nog iets aan 6106 toe 6205*. Door invoering van dit vierde cijfer kunnen in de frequentielijsten de bij elkaar behorende woorden als één woordvorm worden gepresenteerd: *voegde 255 toe 620 en daar 540 aan 610*.

4.3 Enkele opmerkingen bij het codeersysteem C_3C

Het is merkwaardig dat het boek *Woordfrequenties* en het daarin gehanteerde C_3C -systeem in de vakliteratuur amper is besproken. Alleen in de tijdschriften *Onze Taal* en *Meta* zijn commentaren te vinden: *Vervoorn* (1975) en *Gerritsen* (1976). Wel wordt gelukkig steeds meer gebruik gemaakt van gegevens verkregen door woordfrequentie-onderzoek: *Van Heuven* (1978), *Pijnenburg* (1979) en *Gerritsen* (1980). Maar in deze studie, waarin de gegevens uit *Woordfrequenties* zo'n belangrijke plaats innemen en waarin bij het opstellen van hypothesen de mogelijkheden en onmogelijkheden van het C_3C -systeem zo'n doorslaggevende invloed hebben, mogen toch een aantal kleine beperkingen en tekortkomingen van dit systeem niet onvermeld blijven. De opmerkingen hieronder hebben betrekking op de nomenclatuur (4.3.1), op de onderscheidingen die wel en die niet zijn gemaakt, en in samenhang hiermee op de zogenaamde

systeemdwang (4.3.2), en op enkele slordigheden bij de codering (4.3.3). Hierbij zijn niet nogmaals vermeld de in de vorige paragraaf al opgesomde discutabele of merkwaardige beslissingen (bijvoorbeeld het Codeervoorschrift voor het toekennen van code 003 of voor 'bepaald', pagina 52 en 57).

4.3.1 Nomenclatuur

De terminologie bestaat uit een bonte mengeling van Nederlandse en latinistische termen. We vinden 'substantivisch' naast 'zelfstandig', 'adjectivisch' naast 'bijvoeglijk', 'postposities' naast 'achterzetsels' enz. Voor het motief voor dit onderscheid staat in het Codeervoorschrift alleen een aanwijzing bij de algemene opmerkingen over 'Pronomina A': "Voor de nomenclatuur van voornaamwoorden voeren we nu de conventie in dat we de nederlandse termen in hun meer gangbare betekenis zullen gebruiken, en de latijnse wanneer we daarmee de door ons gedefinieerde klassen willen aanduiden." Waarschijnlijk heeft dit motief ook een rol gespeeld bij de naamgeving van de andere onderscheidingen in het codeersysteem. Maar de vraag blijft dan waarom bijvoorbeeld categorie 5.. 'bijwoorden' heet en categorie 7.. 'conjuncties'.

Soms werkt de naamgeving verwarring in de hand. Enkele voorbeelden: Code 540 staat voor 'voornaamwoordelijke bijwoorden', maar uit het Codeervoorschrift blijkt dat het hier gaat om 'aanwijzende en onbepaalde voornaamwoordelijke bijwoorden'. Pas dan wordt de parallelle in de codering voor bijwoorden en voornaamwoordelijke bijwoorden duidelijk: 51. en 54. voor 'aanwijzende en onbepaalde'; 52. en 55. voor 'vragende', en 53. en 56. voor 'betrekkelijke'. Ook bij het derde cijfer voor persoonsvormen is er een onduidelijkheid. De cijfers ..7 en ..8 voor imperativa zijn ingevoerd voor het al dan niet voorkomen van een pronomen personale. Maar uit het Codeervoorschrift blijkt dat het hier gaat om het al dan niet voorkomen van een pronomen personale als subject. Dus ondanks het pronomen personale 'ze' krijgen we toch ..7 in *pak 257 ze!* (Het cijfer ..8 is van toepassing op *weest 278 u flink.*)

De omschrijving van 610 (2e deel gesplitste voornaamwoordelijke bijwoorden) kan eveneens problemen geven, wanneer men het Codeervoorschrift niet kent. Wellicht zal een gebruiker van *Woordfrequenties* denken, dat er een fout is gemaakt wanneer hij ziet staan *in 610*. Want uit de omschrijving van 610 valt af te leiden dat *in 610* alleen maar kan voorkomen als tweede deel van een gesplitst voornaamwoordelijk bijwoord, in combinatie met bijvoorbeeld *waar* als eerste deel. Pas wanneer het

Codeervoorschrift bekend is (zie pagina 64), kan een gebruiker van *Woordfrequenties in 610* interpreteren. Maar wellicht zocht ik te veel achter terminologische eigenaardigheden; zo heb ik tevergeefs geprobeerd te doorgronden waarom er bij het derde cijfer van 'Andere woordsoorten' sprake is van 'overige verbogen vormen' (.3) naast 'andere verbogen vormen' (.6 en .9).

Ondanks deze bezwaren is in deze studie toch de nomenclatuur van C_3C gehandhaafd. Argumenten hiervoor zijn:

1. Door een andere omschrijving te kiezen wordt maar al te gemakkelijk de indruk gewekt dat het over andere grammaticale onderscheidingen gaat, dan zoals gedefinieerd in het Codeervoorschrift.
2. Het veranderen van termen maakt het de lezer lastiger om uitkomsten van deze studie te vergelijken met gegevens uit *Woordfrequenties*, of andere studies die gebruik maken van onderscheidingen uit het C_3C -systeem. Dus, wanneer in deze studie bijvoorbeeld over substantieven wordt gesproken, dan zijn bedoeld de woorden die als substantief gecodeerd worden volgens het Codeervoorschrift.

4.3.2 Onderscheidingen

Na de opmerkingen over de nomenclatuur nu enkele opmerkingen over de onderscheidingen die niet zijn gemaakt. Natuurlijk kan bij elk systeem de vraag worden gesteld: Waarom deze onderscheidingen en geen andere? Maar opvallend is, dat nergens wordt verantwoord waarom sommige onderscheidingen wel en andere niet zijn gemaakt. In het Codeervoorschrift staat op pag. 4: " C_3C is zo opgezet, dat zij naar onze smaak aan bepaalde intuïtieve eisen voldoet." Welnu, over smaak valt niet te twisten. In plaats daarvan hier een paar opmerkingen om aan te geven dat in het C_3C -systeem nog gemakkelijk enkele andere, mijns inziens belangrijke, onderscheidingen kunnen worden aangebracht. Ik wijk hierin af van het oordeel van Brandt Corstius (1978:72), die spreekt over "... een codering (...) die men niet gauw verfiijnder zou wensen."

Bij de hulp- en koppelwerkwoorden hadden naar mijn mening fijnere onderscheidingen aangebracht moeten worden. Waarom bijvoorbeeld geen aparte codering voor de hulpwerkwoorden van de lijdende vorm? In het Codeervoorschrift (pag. 40) wordt opgemerkt dat het onderscheid tussen hulp- en koppelwerkwoorden veelal dubieus is - hetzelfde geldt uiteraard ook voor onderscheidingen binnen de categorie hulpwerkwoorden - maar is dit een reden om van het onderscheid af te zien? Een van mijn "intuïtieve eisen" aan een codeersysteem is, dat op basis daarvan bepaald kan worden hoe vaak van de lijdende vorm gebruik gemaakt wordt.

Zo'n vraag kan nu niet zonder nader onderzoek beantwoord worden. Het argument "dubieus onderscheid" lijkt mij niet terecht: Het onderscheid is er wel, maar het is moeilijk om precieze grenzen te trekken. Hetzelfde geldt voor bepaalde onderscheidingen tussen hulpwerkwoorden en transitief gebruikte werkwoorden, waar de grenzen wel worden afgebakend (Codeervoorschrift, pag. 41-45).

Een ander argument tegen fijnere onderscheidingen zou zijn, dat er binnen het tweede cijfer van categorie 2.. geen ruimte meer is. (Er zijn hier minimaal twee tweede cijfers nodig voor verder onderscheid, omdat er steeds onderscheid wordt gemaakt tussen deelwoorden en infinitieven enerzijds en persoonsvormen anderzijds.) Maar waarom hier een aparte codering voor interjectivisch gebruikte werkwoorden? Als we deze codering weglaten, zijn er wel twee tweede cijfers beschikbaar: 28. en 29. Interjectivisch gebruikte werkwoorden kunnen ook onder 8.., overigens net als 08., 18. en 58. Het bezwaar dat dan niet alle werkwoordsvormen met hetzelfde cijfer beginnen, lijkt mij niet doorslaggevend. Hetzelfde geldt in het C₃C-systeem voor de pronomina (3.. en 4..). En wanneer in een overzicht het percentage werkwoordsvormen genoemd moet worden, kan bijvoorbeeld code 82. voor interjectivisch gebruikte werkwoorden er gemakkelijk bij worden betrokken.

Ook de onderverdeling bij het derde cijfer voor deelwoorden en infinitieven laat te wensen over. Bij de infinitieven wordt er wel onderscheid gemaakt tussen verbaal en niet-verbaal gebruik (in het laatste geval dus 'substantivisch'). Maar bij de deelwoorden gebeurt dit slechts gedeeltelijk. In de volgende twee voorbeelden krijgt *gekozen* dezelfde codering: *de gekozen 216 burgemeester*; *zij hebben hem tot burgemeester gekozen 216*. Hier zou het onderscheid verbaal en niet-verbaal (adjectivisch) zinvol zijn. Ook het onderscheid tussen substantivisch en adjectivisch gebruikte onvoltooide deelwoorden had naar mijn mening aangebracht moeten worden. Nu krijgen we dezelfde codering in: *de volgende 203 opmerkingen* en *in het volgende 203*.

Bij het derde cijfer voor persoonsvormen gelden soortgelijke opmerkingen. Ik had liever meer ruimte gezien voor onderscheidingen naar persoon in het preteritum. Nu 'zijn' alleen gegevens bekend over het preteritum in enkelvoud en meervoud.

Dit alles zou inhouden dat er twee eerste cijfers nodig zijn voor de werkwoordsvormen. De mogelijkheid hiertoe bestaat ook binnen dit driecijfersysteem, want de onderscheidingen bij de categorieën 7.., 8.. en 9.. hadden heel goed binnen één eerste cijfer ondergebracht kunnen worden. Ook uit de volgende voorbeelden blijkt, dat er in dit codeer-

systeem nog ruimte kan worden vrijgemaakt.

Binnen de pronomina wordt steeds onderscheid gemaakt naar zelfstandig en bijvoeglijk gebruik, de personalia uiteraard uitgezonderd. Maar het is zeer de vraag of dit onderscheid bij de relativa wel zinvol is. In het Codeervoorschrift (pag. 58) staat dat *welk(e)* als relativum zo'n interessant woord is "omdat geen ander relativum in het Nederlands ooit bijvoeglijk wordt gebruikt." Dit houdt dus in dat code 43. is ingevoerd voor één woord. Het onderscheid zelfstandig-bijvoeglijk doet hier denken aan systeemdwang. Hetzelfde geldt voor het onderscheid superlatief (genitief), ..8, kennelijk ingegeven door het onderscheid comparatief (genitief), ..5; bijvoorbeeld *iets alleraardigst* 108. De genitief van de superlatief kan in principe voorkomen bij 1., 44., 45. en 5., maar waarom voor zo'n bijzonder geval een aparte plaats openhouden in het systeem? Overigens, het derde cijfer ..8 blijkt niet voor te komen, en het derde cijfer ..5 komt alleen maar voor bij 10. (N.B. Het derde cijfer ..8 bij werkwoordsvormen en 'restgroepen' heeft een andere betekenis.) Was het niet beter geweest om de genitief-onderscheiding bij de superlatief te laten vallen en in plaats daarvan bijvoorbeeld *de laatste* 129 *tijd* en *de laatsten* 129 te onderscheiden? Een laatste voorbeeld waardoor ruimte gecreëerd zou kunnen worden, betreft het tweede cijfer .8. voor interjecties bij 0., 1., 2. en 5.. Ook een vorm van systeemdwang: waar interjecties kunnen worden onderscheiden gebeurt dat, terwijl op voorhand valt aan te nemen dat het aantal interjecties niet bijzonder groot zal zijn.

Tot zover wat betreft andere, wenselijke onderscheidingen bij werkwoorden en de ruimte die daarvoor binnen C₃C gecreëerd zou kunnen worden. Tot slot van deze paragraaf nog een voorbeeld van een andere categorie waarin fijnere onderscheidingen helaas ontbreken, en wel de te vage code 510.

Onder 510 vallen zowel de aanwijzende als de onbepaalde bijwoorden. Er wordt geen argument gegeven voor het samenvoegen van deze twee typen bijwoorden. En het samenvoegen is des te merkwaardiger omdat bij pronomina het onderscheid aanwijzend-onbepaald wel gehanteerd wordt (demonstrativa en indefinita). Men kan tegenwerpen dat het onderscheid tussen 'aanwijzend' en 'onbepaald' heel moeilijk is. Maar er zijn vele andere onderscheidingen gemaakt die eveneens heel moeilijk liggen. Naar mijn mening kan, zeker gezien het Codeervoorschrift voor indefinita, vrij gemakkelijk het onderscheid tussen 'aanwijzend' en 'onbepaald' worden aangebracht. Dit geldt ook voor de bijwoorden die op beide manieren kunnen worden gebruikt. Bijvoorbeeld *daar* in *daar is hij* (aanwijzend) en *daar belt iemand* (onbepaald). Ik beperk me nu tot de

woorden genoemd in het Codeervoorschrift voor 510. Uiteraard kunnen er meer zijn. Onderstreept zijn de bijwoorden die mijns inziens als aanwijzend gecodeerd kunnen worden:

er, daar(?), hier, ergens, overal, nergens, alom, daarginds, ooit, dan, nu, thans, toen, altijd, altoos, immer, nooit, nimmer, soms, dus, hierom, daarom (als het betekent "om diezelfde reden"), zo, zodoende, aldus, dusdoende, zover(re), dusver(re), derwaarts, herwaarts, her, (en) der, zolang, vandaar, daarentegen, zus (in "nu eens zus dan weer zo").

Door deze verfijning aan te brengen - dit zou dan ook moeten gebeuren bij 540, voor de voornaamwoordelijke bijwoorden - kunnen de in *Woordfrequenties* verwerkte teksten gemakkelijker onderzocht worden op het gebruik van onbepaalde woorden (naast de indefinita in elk geval ook de onbepaalde bijwoorden).

Maar er is nog iets anders aan de hand met code 510. Uit het Codeervoorschrift blijkt dat tot deze categorie ook behoren de woorden met een historische genitief of datief of met 'dien'. Woorden als *destijds* of *indertijd* zou ik ook tot de (onbepaalde) bijwoorden rekenen, maar wat te denken van *desondanks*, *toentertijd*, *uitentreure*, *dienovereenkomstig*, *bovendien* enz.? Waarom hier niet code 500?

Dit brengt mij bij de derde opmerking over 510. Op sommige onderdelen is er een zeer arbitraire scheiding met de 'gewone' bijwoorden. De woorden *ooit*, *nooit*, *soms* en *altijd* krijgen 510, maar in het Codeervoorschrift wordt er met nadruk op gewezen dat de, wat ik hier zal noemen 'onbepaalde tijdsaanduidingen': *eens*, *weleer*, *steeds*, *aldoor* enz. code 500 krijgen. De zin van dit onderscheid ontgaat mij. Ook *heden*, *ongeveer*, *onlangs*, *vaak* en *voorheen* krijgen code 500. Waarom hier niet 510? En *momenteel*, mijns inziens toch een aanwijzend bijwoord, krijgt zelfs code 150. Kortom, code 510 blijft te vaag.

4.3.3 Slordigheden

Na de kanttekeningen bij de nomenclatuur en de onderscheidingen binnen het C₃C-systeem tot slot nog enkele opmerkingen over de slordigheden in de codering. Bij dit toch vrij ingewikkelde codeerwerk zijn fouten niet te vermijden. Vooral wanneer je het codeersysteem goed kent, vallen fouten op als: *betrokkene* 216 in A₂, *enkele* 450 naast *enkele* 453 in A₁, *er* 540 in A₁ en *er* 516 in A₂, (*mee* 620) *brengt* 210 in A₂ en *ze* 330 in A₁. (Met de aanduidingen A₁, A₂ enz. worden frequentielijsten bedoeld; zie hierover de volgende paragraaf.)

Dergelijke fouten zijn moeilijk te vermijden. Maar dit geldt niet voor andere slordigheden: Het pronomen *meer* moet volgens het Codeervoorschrift altijd als comparatief gecodeerd worden, dus 444 of 454. Maar toch vinden we in A_1 *meer 440* en in A_2 *Meer 440*.

Een aardig voorbeeld van een woord dat verschillende codes kan krijgen, is *verdergaande*. Moeten we dit woord opvatten als een verbogen adjectief (103), als een onverbogen comparatief (104), als verbogen comparatief (106), of als verbogen onvoltooid deelwoord (203)? Alle vier codes komen voor in A_2 . Bij dit probleem laat het Codeervoorschrift ons in de steek. Maar waarom hier geen knoop doorgehakt? Uiteraard kan *verdergaande* ook voorkomen als bijwoordelijk gebruikt voltooid deelwoord (205), maar het onderscheid tussen deelwoorden en bijwoordelijk gebruikte deelwoorden wordt in het Codeervoorschrift wel verantwoord. Vergelijk voor andere twijfelgevallen nog: *au serieux* in A_2 met code 500 (bijwoord) en met code 990 ('anderstalig'), of *gaarne 500* en *graag 150*.

Nog een andere slordigheid had gemakkelijk vermeden kunnen worden, namelijk een code binnen een woordgroep die volgens het Codeervoorschrift als één woord had moeten worden gecodeerd. Het betreft hier vooral eigennamen. Zo vinden we in A_3 *Provinciale Staten 001* maar ook *Provinciale 001 Staten 001*, en verder onder andere *Westfries 010 Museum 010* in plaats van *Westfries Museum 010* en merkwaardigerwijs ook *van 010 Journalisten 010* en *voor 010 Dialectologie 010*. In elke kolom uit A_3 komen we wel een paar van dergelijke voorbeelden tegen. Dit heeft tot gevolg dat de woorden 'Westfries', 'van' en 'voor' als aparte eigennamen zijn geteld bij het opstellen van frequentielijsten van codes. De in *Woordfrequenties* opgegeven frequentie van 010 is dus iets hoger dan in werkelijkheid. Dit had men door extra controle gemakkelijk kunnen voorkomen.

Tot slot nog een opmerking over woorden met een kennelijk foutieve code. Deze fouten zijn vooral in A_2 te vinden. Wanneer we bijvoorbeeld de vijf meest frequente woorden uit het totaal corpus 'geschreven taal' nemen, *de 370*, *van 600*, *een 450*, *en 700*, *het 370*, en deze woorden opzoeken in A_2 , dan vinden we de volgende, kennelijk foutieve, codes: *de 103*, *de 360*, *de 600*; *van 000*; *een 700*, *een 710*; *en 300*; *het 450*, *het 600*, *het 990*. Ook in A_1 komen dit soort fouten voor. Daar zien we bijvoorbeeld dat *een* met code 998 vijf maal voorkomt in gezinsbladen. Maar code 998 wordt gebruikt voor 'niet tot de steekproef behorende elementen' (zie pagina 68) en is ook niet opgenomen in de frequentielijsten van codes. Deze fouten kunnen gemakkelijk ontstaan wanneer bij

het samenvoegen van woorden en bijbehorende codes een woord wordt overgeslagen. (Zie hierover paragraaf 5.4, Tekstbewerking.) Maar deze fouten zijn ook gemakkelijk op te sporen wanneer een lijst van woorden, alfabetisch gerangschikt naar code, wordt gecontroleerd.

4.4 De frequentielijsten

In *Woordfrequenties* staan negen frequentielijsten met gegevens per subcorpus (dagbladen, opiniebladen, gezinsbladen enz.) en per deeltotaal ('totaal geschreven taal' en 'totaal gesproken taal').

In de frequentielijsten A_1 , A_2 en A_3 staan de woordvormen alfabetisch geordend. Woordvormen wil hier ook zeggen dat vormen met hoofdletters, accenttekens en afbrekingstekens apart zijn opgenomen. De A-lijsten zijn als volgt te onderscheiden:

- A_1 woordvormen, niet langer dan 25 tekens, die meer dan één keer voorkomen;
- A_2 woordvormen, niet langer dan 25 tekens, die slechts één keer voorkomen;
- A_3 woordvormen die langer zijn dan 25 tekens.

Bij 'teken' moet bedacht worden dat een hoofdletter als twee tekens geldt, dat een spatie ook een teken is en dat een code zelf al drie tekens bevat. Dus de woordvorm *Buit- tenlandse Zaken 001* (26 tekens) hoort thuis in A_3 . Maar *Buitenlandse Zaken 001* (24 tekens) hoort thuis in A_1 of A_2 . Het onderscheid tussen A_1 , A_2 en A_3 wordt nergens verantwoord. Maar in elk geval wordt er zo wel veel papier bespaard. Wanneer de woordvormen worden afgedrukt met frequenties gespecificeerd naar subcorpora, dan nemen de woordvormen die slechts één keer voorkomen of lang zijn, veel ruimte in beslag. Door de lange woordvormen apart te nemen in A_3 konden de woordvormen die meer dan één keer voorkomen, in twee kolommen worden afgedrukt. De woordvormen die maar één keer voorkomen (A_2) konden in deze opzet in vijf kolommen worden gepresenteerd. (Achter de woordvormen hoeft dan slechts één aanduiding voor het subcorpus te staan.)

De B-lijst is een alfabetische frequentielijst van lemmata. Hier zijn bijvoorbeeld alle vormen van het werkwoord 'lopen' bij elkaar opgeteld. In de A-lijsten vinden we elke woordvorm met een drie-cijfercodering, maar in de B-lijst de lemmata met, uiteraard, alleen het eerste codecijfer. Van belang is nog, dat niet alle woordvormen uit de A-lijsten in de B-lijst staan vermeld. Alleen die lemmata zijn opgenomen waarvan de totaalfrequentie tenminste 5 bedraagt.

In de C-lijst staan alle woordvormen gerangschikt naar dalende frequentie. Uit deze lijst kunnen we bijvoorbeeld halen wat de tien meest frequente substantieven in gezinsbladen zijn.

De drie D-lijsten bevatten informatie over de frequenties van woordklassen:

- D₁ woordklassen, geordend naar lexicale code, geteld naar tokens;
- D₂ woordklassen, geordend naar lexicale code, geteld naar types;
- D₃ percentuele distributie van woordklassen.

In D₁ staat achter elke code hoe vaak die voorkomt. D₂ geeft aan hoeveel verschillende woordvormen een zelfde code krijgen. Wanneer we bijvoorbeeld een aparte code zouden hebben voor de kleine categorie lidwoorden, dan zou er in D₂ een heel laag getal staan achter deze code, maar in D₁ een hoog getal omdat die paar lidwoorden heel vaak voorkomen.

In D₃ staat van een aantal grammaticale categorieën vermeld hoe ze percentueel in subcategorieën zijn onderverdeeld. Uit deze lijst kunnen we bijvoorbeeld halen hoe groot het percentage werkwoordsvormen is in dagbladen, en hoe groot daarbinnen het percentage intransitief gebruikte werkwoorden is.

Lijst E, tot slot, geeft informatie over frequentieklassen. Dat wil zeggen dat we uit deze lijst kunnen halen hoeveel woordvormen een frequentie hebben van 1, of bijvoorbeeld een frequentie tussen 300 en 309.

4.5 Enkele opmerkingen bij de frequentielijsten

In dit onderzoek is vooral gebruik gemaakt van gegevens uit de A- en D-lijsten. Maar helaas wordt de bruikbaarheid van deze lijsten enigszins beperkt door de manier waarop ze zijn gepresenteerd.

In de A-lijsten zijn, zoals vermeld, woordvormen afgedrukt al dan niet met hoofdletters, accenttekens of afbrekingstekens. Maar ik betwijfel sterk de waarde van deze onderscheidingen. In de inleiding op *Woordfrequenties* wordt overigens, naar aanleiding van de informatie over hoofdletters en afbrekingsstreepjes, ook al vermeld dat de gebruiksmogelijkheden van dit soort onderscheidingen tamelijk beperkt zijn. Hoe het ook zij, een gevolg van deze manier van presenteren is, dat voor bepaalde vragen de A-lijsten minder gemakkelijk zijn te raadplegen. Een voorbeeld: wanneer we de frequentie opzoeken van *Nederlandse* 103, vinden wij in A₁ vier woordvormen (*Nederlandse*, *Ne-derlandse*, *Neder-landse* en *Nederland-se*). We moeten dan dus zelf de frequenties van

deze woordvormen bij elkaar optellen. Het zal duidelijk zijn dat we voor deze vraag niet in de B-lijst, de alfabetische frequentielijst van lemmata, terecht kunnen; daar vinden wel alleen het lemma *Nederlands* 1. En gesteld dat we de frequentie van een lemma willen weten, dan nog zijn we gedwongen bij lemmata waarvan de woordvormen een frequentie van minder dan 5 hebben, de A-lijsten op te slaan.

Maar ook de A_1 -lijst geeft geen afdoend antwoord op onze vraag. Want het is mogelijk dat er nog een woordvorm 'Néderlandse' is die maar één keer voorkomt; deze vorm is dan ondergebracht in lijst A_2 . Door de presentatie in woordvormen zijn we dus altijd gedwongen de A_1 - en de A_2 -lijst op te slaan om antwoord te krijgen op vragen over frequenties van woordvormen.

Door de manier van presenteren zijn we ook verplicht om de A_2 - en de A_1 -lijst te raadplegen, wanneer we informatie willen hebben over woorden die slechts één keer voorkomen. Want in A_2 staan woorden, weliswaar in een andere woordvorm, die ook in A_1 zijn te vinden. Ter illustratie enkele gegevens uit een willekeurige kolom in A_2 ; elke kolom bevat 80 woordvormen. In de kolom die begint met 'kook-boek' staan acht woordvormen die in een andere woordvorm ook voorkomen in A_1 , bijvoorbeeld *koppe-ling* in A_2 naast *koppeling* in A_1 . (In dezelfde kolom komen overigens nog drie woordvormen voor die in een andere woordvorm ook één keer voorkomen in hetzelfde subcorpus.) A_2 geeft dus zonder A_1 geen informatie over woorden die maar één keer voorkomen, terwijl dit gemakkelijk had gekund, wanneer men had afgezien van de presentatie in woordvormen.

De presentatie in woordvormen, al dan niet met hoofdletters, accenttekens of afbrekingstekens, en trouwens ook de in 4.3.3 vermelde slordigheden, hebben ook invloed op de D_2 -lijst, de lijst waarin vermeld staat hoeveel types per code voorkomen. Een voorbeeld. Op pagina 59 staat bij relativa als bijzonderheid al genoteerd, dat er maar één relativum is dat bijvoeglijk gebruikt kan worden, namelijk *welk(e)*. Op grond hiervan zou men in D_2 bij 430 en bij 433 het cijfer 1 verwachten. Maar in D_2 staat onder 'opiniebladen' een 2 bij 433. Deze 2 is in dit geval het gevolg van het voorkomen van nog een andere woordvorm, namelijk *Welke*. Deze informatie is te halen uit de niet-gepubliceerde lijst van woorden, alfabetisch gerangschikt naar code.

De D_2 -lijst is dus niet te gebruiken wanneer we willen weten hoeveel verschillende woorden per code voorkomen. Bij het raadplegen van de D_2 -lijst moet altijd bedacht worden, dat een type gerelateerd is aan een woordvorm en niet aan een woord. Bovendien kan het aantal typen in

werkelijkheid kleiner zijn door de hierboven vermelde slordigheden in vooral A_2 .

Ook de D_3 -lijst laat, wat presentatie betreft, te wensen over. Deze lijst is in veel gevallen te vaag. Bij het stellen van vragen over percentuele distributies van woordklassen moet steeds nauwkeurig nagegaan worden, hoe het codeersysteem in elkaar zit en hoe het Codeer-voorschrift luidt. Wanneer we bijvoorbeeld willen weten hoe hoog het percentage zelfstandige naamwoorden is in een bepaalde subcorpus, dan kan niet volstaan worden met het noteren van het percentage onder substantieven (0..). Het overzicht van het C_3C -systeem leert ons, dat code 12. wordt gehanteerd voor substantivisch gebruikte adjectieven, *de jongere, de goeden* enz., en dat code 2.1 wordt gehanteerd voor substantivische infinitieven, *het jagen*. Ook code 81. wordt gebruikt voor bepaalde substantieven. Maar dit niet alleen, uit de toelichting op het codeersysteem blijkt, dat ook achter andere codes zelfstandige naamwoorden zitten verscholen; bijvoorbeeld de codes 2.4 en 2.8, deelwoorden in het meervoud: *studerenden, ondergetekenden*. En dan nog zijn we er niet, want achter weer andere codes kan een zelfstandig naamwoord verscholen zitten: de codes 2.3 en 2.7, de verbogen deelwoorden in *het volgende* en *de getroffene*.

De D_3 -lijst geeft een aardig algemeen overzicht, maar voor het nauwkeurig beantwoorden van vragen kunnen beter de frequenties achter de codes uit de D_1 -lijst worden bekeken. En ook dan nog komen we soms niet tot een nauwkeurig antwoord. In dit geval bijvoorbeeld stuiten we bij de vraag naar het percentage zelfstandige naamwoorden op de moeilijkheid dat *volgende* in *het volgende* en *de volgende keer* dezelfde code krijgt.

In hoofdstuk 6, Hypotheses en Toetsing, zullen deze en soortgelijke problemen nader worden besproken. Na deze uiteenzetting over de mogelijkheden en beperkingen van het in deze studie gekozen codeersysteem, volgen nu eerst enkele gegevens over de steekproef uit de taal van 'Den Haag' waarop het onderzoek is verricht.

5 DE STEEKPROEF, SELECTIE EN BEWERKING

5.1 Inleiding

In dit hoofdstuk staat beschreven hoe het onderzoeksmateriaal is geselecteerd en hoe het is bewerkt. Of met andere woorden: hoe eerst een steekproef is getrokken van bijna 500 fragmenten - dit wil zeggen bijna 2000 zinnen of ongeveer 50.000 woorden - uit ongeveer 22.000 pagina's correspondentie tussen regering en Staten-Generaal, en hoe daarna de woorden zijn gecodeerd en de steekproef is bewerkt voor het toetsen van hypothesen, zoals die zijn af te leiden uit de oordelen vermeld in hoofdstuk 3.

In deze inleiding komt eerst het belang van een nauwkeurige verslaggeving van de steekproeftrekking (5.1.1) ter sprake. Daarna het bronnenmateriaal waaruit de steekproef is getrokken (5.1.2). Vervolgens enkele bezwaren tegen het onderzoeksmateriaal (5.1.3). En tot slot enkele opmerkingen over de omvang van de steekproef (5.1.4).

Paragraaf 5.2 is gewijd aan een beschrijving van de manier waarop de steekproef is getrokken en de arbitraire beslissingen die hierbij zijn genomen. De steekproeftrekking is in twee fasen verlopen: het selecteren van pagina's en het selecteren van fragmenten uit die pagina's. Hierbij is aangegeven welke pagina's en welke tekstgedeelten van eenmaal geselecteerde pagina's uit het bronnenmateriaal buiten beschouwing zijn gelaten, dus uit welke populatie de steekproef uiteindelijk is getrokken. In paragraaf 5.3 staat een overzicht van de herkomst van de onderzochte fragmenten met enkele voorbeelden. Tot slot wordt in 5.4, Tekstbewerking, beschreven hoe de codeerwerkzaamheden zijn uitgevoerd en welke andere bewerkingen zijn toegepast op de steekproef.

5.1.1 Belang van de verslaggeving

Als algemene waarheid geldt, dat een nauwkeurige beschrijving van de manier waarop onderzoeksmateriaal wordt geselecteerd van groot belang is voor het beoordelen en eventueel generaliseren van onderzoeksresultaten. Immers, pas bij een nauwkeurige beschrijving kan gecontroleerd worden of - en zo ja, hoe - de materiaalselectie de onderzoeksresultaten heeft beïnvloed. Voor dit onderzoek naar de eventuele eigenaar-

digheden van de taal van 'Den Haag' in vergelijking met andere vormen van taalgebruik kan het volgende voorbeeld deze algemene waarheid verduidelijken.

Wagner (1970) heeft een frequentie-onderzoek uitgevoerd op 1000 zinnen uit teksten van de overheid, en die zinnen vergeleken met journalistentaal. Maar in dit onderzoek staat niet vermeld hoe de steekproef is getrokken. Bovendien is niet aangegeven welke soorten overheidsteksten of welke vormen van journalistiek taalgebruik onderdeel hebben uitmaakt van het onderzoek. Een van de opvallendste conclusies van Wagner is, dat de korte zin kenmerkend is voor overheidsteksten (a.w.:45).

Doch deze conclusie is van weinig belang omdat de manier waarop de steekproef tot stand is gekomen, van invloed kan zijn op de gemiddelde zinslengte van de steekproef. Ook Spang-Hanssen (1963:63-66) wees hier al op, onder verwijzing naar Yule (1939:382): "measurement of sentence lengths is a field involving particular dangers of bias due to inappropriate sampling." Hij geeft dan een voorbeeld, dat hier voor een deel en vereenvoudigd wordt weergegeven.

Stel dat een tekst afwisselend zinnen bevat van 20 en van 40 woorden, dan is de gemiddelde zinslengte van die tekst 30 woorden. Maar als de steekproef wordt getrokken door bijvoorbeeld uit elke pagina de zin te nemen waarin de tiende regel zit, komen er naar verhouding meer lange zinnen in de steekproef. Want een zin van 40 woorden loopt over meer regels dan een zin van 20 woorden. En als de steekproef tot stand komt door bepaalde regels te prikken, dan hebben de zinnen van 40 woorden meer kans om in de steekproef terecht te komen. Deze kans valt gemakkelijk te berekenen. Het resultaat op basis van deze steekproeftrekking zou dan zijn een gemiddelde zinslengte van 33,3 woorden.

Ook de Werkgroep Frequentie-Onderzoek heeft een selectiemethode toegepast die de zinslengte van de steekproef beïnvloedt (zie het Verslag van de Werkgroep, 1974:19-28). Voor het corpus 'dagbladen' is men begonnen met een tekstgedeelte uit de eerste kolom van de eerste pagina van een krant. Vervolgens heeft men uit de tweede kolom van de tweede pagina een tekstgedeelte genomen, dat begon op de hoogte waar het tekstgedeelte uit de eerste pagina ophield. Op deze manier heeft men een grotere kans om lange zinnen te selecteren. Voor de andere corpora heeft de Werkgroep Frequentie-Onderzoek een soortgelijke procedure gevolgd. Daarom zijn de in deze studie vermelde onderzoeksresultaten over zinslengte niet direct te vergelijken met zinslengte-gegevens van de corpora uit *Woordfrequenties*. (Zie hierover pagina 145 - 147.) Met dit voorbeeld over een nog relatief eenvoudig verschijnsel als

zinslengte is voldoende aangetoond dat het bijzonder belangrijk is, precies aan te geven hoe het onderzoeksmateriaal is geselecteerd.

5.1.2 Het bronnenmateriaal

De onderzochte tekstfragmenten zijn geselecteerd uit de correspondentie tussen regering en Staten-Generaal in het parlementaire jaar 1975-1976 (16 september 1975 - 21 september 1976). De stukken die door regering en Staten-Generaal worden gewisseld, verschijnen voornamelijk als Bijlagen en voor een klein deel als Aanhangsel bij de Handelingen. De Handelingen zelf, het stenografisch verslag van de vergaderingen van de Eerste en Tweede Kamer der Staten-Generaal, zijn niet in het onderzoek betrokken; deze studie is beperkt tot geschreven taal. De eigenaardigheden in schriftelijke weergave van gesproken taal kunnen vertroebelend werken.

In het Aanhangsel staan de door kamerleden schriftelijk gestelde vragen met de daarop door de regering gegeven antwoorden. In het parlementaire jaar 1975-1976 zijn er 1682 vragen gesteld en beantwoord. Bijna alle vragen waren gesplitst in een aantal samenhangende 'deelvragen'. Het Aanhangsel over 1975-1976 nam, volgens een opgave van de afdeling Officiële Publikaties van de Staatsuitgeverij, 2188 bedrukte pagina's in beslag.

De Bijlagen zijn veel moeilijker te karakteriseren: ze bevatten alle andere stukken die voor een van beide Kamers zijn gedrukt. Op het tijdstip dat dit onderzoek is gestart (direct na Prinsjesdag 1976) waren over de Bijlagen 1975-1976 heel weinig gegevens bekend. De totale omvang bedroeg over het kalenderjaar 1975 "53 miljoen tekens op 20.011 pagina's", aldus het Jaarverslag over 1975 van het Staatsdrukkerij- en Uitgeverijbedrijf (pag. 21). En verder was het aantal moties en amendementen bekend; volgens een opgave van de griffie van de Tweede Kamer respectievelijk 387 en 755. Dit gegeven komt straks nog aan de orde, bij de selectieprocedure voor de pagina's (5.2.2), om het verschil aan te geven tussen een aselechte en een representatieve steekproef. De nu volgende korte beschrijving van de Bijlagen heeft als doel enig inzicht te geven in het belangrijkste deel van het bronnenmateriaal.

De Bijlagen - in de wandeling doorgaans kamerstukken genoemd - bestaan voor verreweg het grootste deel uit de volgende categorieën tekstep:

1. teksten die betrekking hebben op het voorbereidend onderzoek voor de openbare behandeling van wetsontwerpen:
 - wetsontwerpen en memories van toelichting daarop van de regering;

- verslagen (onder andere: voorlopige verslagen, eindverslagen en verslagen van een mondeling overleg) van commissies uit de Eerste en Tweede Kamer;

- memories van antwoord en nota's naar aanleiding van een verslag van de regering.

2. teksten die tijdens de plenaire beraadslaging in een van beide Kamers ter tafel komen:

- amendementen, nota's van wijziging en moties.

3. teksten die de regering aan de Eerste en Tweede Kamer voorlegt en die niet direct te maken hebben met de wetgeving:

- stukken die dienen ter voorbereiding van de openbare behandeling van nota's en brieven;

- nota's;

- brieven.

Hierbij zij nog opgemerkt dat categorie 2 verreweg het kleinst is en dat de categorieën 1 en 3 elkaar in omvang niet veel ontkenen. De kamerstukken worden door de griffies van de Eerste en Tweede Kamer voorzien van een hoofdnummer en een volgnummer. Alle stukken krijgen een hoofdnummer per onderwerp volgens datum van verschijnen. De stukken uit het parlementaire jaar 1975-1976 zijn verdeeld over de hoofdnummers 13.600 tot 14.100. De stukken die op hetzelfde onderwerp betrekking hebben, krijgen een volgnummer, ook naar datum van verschijnen. Een voorbeeld: in de correspondentie over een wetsontwerp kunnen als volgnummers voorkomen: 1. koninklijke boodschap; 2. ontwerp van wet; 3. memorie van toelichting; 4. voorlopig verslag; 5. memorie van antwoord; 6. eindverslag; 7. nota naar aanleiding van het eindverslag; 8-12 amendementen; 13-17 moties. Maar het aantal volgnummers per hoofdnummer verschilt sterk; het kan ook zijn: 1. geleidende brief; 2. nota.

Helaas is er niets te zeggen over het aantal pagina's per hoofdnummer (onderwerp) en volgnummer (soort kamerstuk), want de kamerstukken worden alleen per volgnummer gepagineerd. Daarom was het heel moeilijk om een steekproef te trekken waarin de verschillende onderwerpen of verschillende soorten kamerstukken evenredig zijn vertegenwoordigd.

5.1.3 Bezwaren tegen het onderzoeksmateriaal

Tegen de keuze van het onderzoeksmateriaal, de correspondentie tussen regering en Staten-Generaal uit het parlementaire jaar 1975-1976, zijn vier bezwaren aan te voeren, die alle te maken hebben met de vraag:

Is deze correspondentie wel geschikt als materiaal voor het toetsen van hypothesen over de taal van 'Den Haag'? Deze bezwaren zijn geen van alle afdoende te weerleggen. Hoogstens kan aannemelijk worden gemaakt dat deze correspondentie, of beter de gedeelten daaruit die in de steekproef zijn opgenomen, geschikt onderzoeksmateriaal vormt. Bovendien kunnen de praktische problemen genoemd worden die zich voordoen bij het selecteren van ander onderzoeksmateriaal. Hieronder volgen de bezwaren met commentaar.

Bezwaar 1. Is deze correspondentie wel representatief voor de taal van 'Den Haag'? Waarom geen andere teksten genomen? Bijvoorbeeld brieven van gemeentelijke instellingen, circulaires en voorlichtingsbrochures, antwoorden op verzoekschriften van burgers, of publikaties die zijn opgenomen in de 'Bibliografie van in Nederland verschenen officiële en semi-officiële publikaties'?

Commentaar. De beperking tot deze correspondentie is vooral ingegeven door praktische motieven. Deze correspondentie - een belangrijke uitgave van de Staatsuitgeverij, het bedrijf waar ik aan het begin van het onderzoek werkzaam was - is gemakkelijk te verzamelen. Bovendien vormt dit materiaal een eenheid waaruit betrekkelijk eenvoudig een verantwoorde selectie kan worden gemaakt.

Bij alle suggesties voor ander onderzoeksmateriaal moet het volgende opgemerkt worden: Het is bijzonder moeilijk, een verantwoorde selectie te maken uit andere publikaties van de overheid, bijvoorbeeld gemeentelijke circulaires of, heel ruim, officiële publikaties. Gesteld bijvoorbeeld dat dit onderzoek was beperkt tot teksten die de Commissie Duidelijke Taal onder ogen kreeg (zie pagina 1), dan zou men kunnen tegenwerpen dat deze teksten nog niet representatief behoeven te zijn voor de taal van 'Den Haag'. Ook bij elke andere steekproef, uit alle soorten overheidsteksten, kan de vraag gesteld worden: Is de steekproef wel representatief?

Uiteraard is hiermee de vraag naar de representativiteit niet beantwoord. In dit verband kan nog wel worden opgemerkt dat de correspondentie tussen regering en Staten-Generaal een zekere status heeft binnen het grote geheel van overheidsteksten, en dat het daarom in de rede ligt, te veronderstellen dat het taalgebruik in andere overheidsteksten niet in belangrijke mate zal afwijken van het taalgebruik in deze correspondentie.

Bezwaar 2. De correspondentie tussen regering en Staten-Generaal bestaat uit verschillende soorten teksten: wetsteksten, stukken waarin het politiek taalgebruik de boventoon voert en teksten van ambtenaren. Is het

wel juist om juridisch, politiek en ambtelijk taalgebruik in één steekproef op te nemen?

Commentaar. Het vermoeden lijkt gewettigd dat er onderscheid bestaat tussen ambtelijk, juridisch en politiek taalgebruik. (Ik druk mij hier voorzichtig uit omdat het onderscheid nog niet is aangetoond.) Inderdaad is het niet zinvol om drie verschillende (?) typen taalgebruik in één steekproef onder te brengen, daar ze immers door onderling verschillende factoren beïnvloed kunnen worden.

Deze steekproef is dan ook beperkt tot wat hier gemakshalve 'ambtelijke teksten' wordt genoemd. Politiek taalgebruik (de Handelingen van de Eerste en Tweede Kamer der Staten-Generaal) en juridisch taalgebruik (bijvoorbeeld wetsteksten) zijn zoveel mogelijk buiten beschouwing gelaten. Overigens is het lang niet altijd eenvoudig, een strikt onderscheid te maken tussen ambtelijk, juridisch en politiek taalgebruik. Beslissingen hierover blijven tot op grote hoogte arbitrair. Vragen van kamerleden aan de regering zijn in de steekproef opgenomen, hoewel die vragen toch zijn opgesteld door politici of hun medewerkers, evenals teksten naar aanleiding van een wetsontwerp, hoewel er veel voor te zeggen valt deze als juridische teksten te beschouwen.

Het probleem van het onderscheid tussen ambtelijk, juridisch en politiek taalgebruik zou zich echter ook hebben voorgedaan bij een andere steekproef. In bijvoorbeeld een invulformulier of een voorlichtingsbrochure kunnen eveneens politiek of juridisch getinte passages voorkomen. Ook dan moet van te voren besloten worden op grond van arbitraire en praktisch toepasbare onderscheidingen, welke tekstgedeelten buiten beschouwing moeten blijven als het gaat om ambtelijk, en niet om politiek of juridisch taalgebruik.

Bezwaar 3. De oplage van kamerstukken en regeringsnota's is vrij klein. Is het dan niet beter om teksten te onderzoeken die een grotere oplage hebben?

Commentaar. De oplage van de correspondentie tussen regering en Staten-Generaal is inderdaad niet groot. In het Jaarverslag Staatsdrukkerij- en Uitgeverijbedrijf 1976:20 staan de volgende gegevens: Het aantal abonnees bedroeg 1328, met daarnaast een behoorlijke omzet aan losse kamerstukken of nota's en rapporten in boekvorm. Verder waren er eind 1976 5585 abonnementen op gedeelten van de correspondentie. Maar deze betrekkelijk kleine oplage is nog geen reden om de correspondentie niet als onderzoeksmateriaal te gebruiken, want kamerstukken en regeringsnota's hebben door hun status invloed op andere overheidsteksten.

Bezwaar 4. De teksten die zijn verwerkt in *Woordfrequenties* en waarmee de correspondentie wordt vergeleken, zijn afkomstig uit de jaren 1965-1971 (zie het Verslag Werkgroep Frequentie-Onderzoek 1974:14-16). Is het zinvol teksten uit 1975-1976 te vergelijken met teksten van vijf tot tien jaar daarvoor, gezien het feit dat het taalgebruik vermoedelijk verandert?

Commentaar. Op deze vraag valt eigenlijk geen zinnig antwoord te geven: Er is niets bekend over de snelheid waarmee taalgebruik verandert. Wel lijkt het waarschijnlijk dat de ene vorm van taalgebruik (bijvoorbeeld reclametaal) sneller verandert dan de andere vorm (bijvoorbeeld wetenschappelijk taalgebruik). In dit verband zij nogmaals gewezen op de in paragraaf 3.1.1 aangehaalde publikatie van *Vervoorn* (1978) waaruit blijkt dat de kritiek van Van Lennep in 1844 op de taal van de overheid niet veel verschilt van de hedendaagse kritiek.

Overigens moet nog opgemerkt worden dat vier van de vijf subcorpora uit *Woordfrequenties* afkomstig zijn uit de jaren 1969-1971. Alleen het subcorpus populair-wetenschappelijke boeken komt uit de periode 1965-1971. De afwijkende lengte van deze periode wordt (zie het Verslag van de Werkgroep 1974:15) "mede gerechtvaardigd doordat, naar men kan veronderstellen, het taalgebruik van de wetenschap minder dan andere soorten van taalgebruik, onderhevig is aan verandering."

Ik had natuurlijk dit argument kunnen overnemen voor het taalgebruik van de overheid en bijvoorbeeld vijf verschillende jaargangen correspondentie kunnen betrekken in mijn steekproef. Maar hiervan is afgezien om twee redenen.

Ten eerste: Er zijn weinig gegevens over de opbouw van een jaargang (zie pagina 81). Bovendien zijn de verschillende stukken binnen één jaargang niet doorlopend genummerd. Daarom was ik gedwongen (zie paragraaf 5.2.2) de ongeveer 20.000 pagina's uit één jaargang eigenhandig te tellen. Eén jaargang is dan al meer dan genoeg. Overigens zou de hypothese dat ook de taal van de overheid langzaam verandert, omgekeerd juist een reden kunnen zijn om slechts één jaargang correspondentie te onderzoeken.

Ten tweede: Als er teksten van vóór 1975 waren genomen, had men kunnen tegenwerpen dat omstreeks dat jaar de kritiek op de taal van 'Den Haag' weer aardig was toegenomen (zie onder andere het Verslag van de Commissie Duidelijke Taal uit 1974), en dat daarom bevindingen op basis van teksten van voor die tijd geen opgeld meer zouden doen, omdat het taalgebruik van de overheid door deze kritiek misschien wel was veranderd.

5.1.4 De omvang van de steekproef

De omvang van de steekproef bedraagt ruim 48.000 woorden. Om duidelijk te maken hoe deze omvang tot stand is gekomen, is enige informatie vereist over de oorspronkelijke en de gewijzigde opzet van dit onderzoek. Aanvankelijk was het doel, steekproeven te trekken uit de volgende categorieën overheidsteksten:

1. Correspondentie tussen regering en Staten-Generaal.
2. Publikaties die zijn opgenomen in de 'Bibliografie van in Nederland verschenen officiële en semi-officiële publikaties', en dan vooral rapporten van ambtelijke commissies en overheidsinstellingen buiten de departementen, om een overlapping met categorie 1 te voorkomen;
3. Brieven van gemeentelijke instellingen, circulaires en voorlichtingsbrochures, antwoorden op verzoekschriften van burgers, enz.

Uit elk van deze groepen zou een steekproef getrokken worden van ongeveer 40.000 woorden zodat de drie steekproeven samen even groot zouden zijn als één van de zes subcorpora uit *Woordfrequenties*, namelijk 120.000. Het getal 40.000 is dus volstrekt arbitrair. Trouwens, ook de Werkgroep Frequentie-Onderzoek heeft de omvang van haar corpus arbitrair vastgesteld. Zie hierover het Verslag van de Werkgroep (1974:11):

"De totale omvang van het Corpus (dit is het materiaal dat de werkgroep heeft laten onderzoeken) bedraagt ongeveer 720.000 woorden. Dit getal is enigszins geïnspireerd op de omvang van het zgn. Brown Corpus voor hedendaags geschreven Amerikaans Engels. Dit bedraagt weliswaar ca. 1.000.000 woorden. Maar aanvankelijk was het de bedoeling in ons Corpus evenveel gesproken als geschreven taalmateriaal op te nemen, nl. 8 subcorpora (van beide soorten 4) van ca. 125.000 woorden. Wegens de specifieke moeilijkheden met het opzetten van een Corpus Gesproken Taal is dit voornemen gewijzigd: er werden 6 subcorpora gecreëerd van ongeveer 120.000 woorden, waarvan het subcorpus Gesproken Taal er een is." Voor de omvang van de subcorpora geeft de Werkgroep dus eigenlijk alleen als argumentatie: het getal 1.000.000 uit het zogenaamde Brown Corpus en het plan voor de 8 subcorpora. En verder is deze aanvankelijke doelstelling gewijzigd vanwege de hierboven aangehaalde niet nader gespecificeerde 'specifieke moeilijkheden'.

De oorspronkelijke opzet van dit onderzoek is eveneens gewijzigd. Na de selectie- en codeerwerkzaamheden van de eerste steekproef, waarmee alvast een begin was gemaakt, is besloten alleen de onder nummer 1 genoemde steekproef in het onderzoek te betrekken, en wel om de volgende redenen:

1. De eerste steekproef en de vergelijkingsmogelijkheden met het materiaal uit *Woordfrequenties* leverden al zoveel studiemateriaal op, dat een grondige bestudering hiervan beter en prettiger was dan nog meer, nauwelijks taalkundig te noemen en bovendien tijdrovende, selectie- en codeerwerkzaamheden uit te voeren.

2. Programmeerarbeid en computerseconden zijn kostbaar.

Overigens bestaat deze steekproef uit meer dan 40.000 woorden, namelijk 48.257; dit komt door een foute schatting bij het selecteren van de fragmenten (zie pagina 92).

5.2 De steekproef

In deze paragraaf staat beschreven hoe het onderzoeksmateriaal is geselecteerd uit de ruim 20.000 pagina's Bijlagen en de 2188 pagina's Aanhangsel. De materiaalselectie verliep in twee fasen: 1. het selecteren van de pagina's; 2. het selecteren van fragmenten uit die pagina's. Bij het samenstellen van de steekproef zijn drie arbitraire beslissingen genomen. Hierover eerst.

5.2.1 Drie arbitraire beslissingen

De doelstelling wat de omvang betreft, 40.000 woorden, is als volgt geconcretiseerd: Selecteer uit de ruim 20.000 pagina's Bijlagen en de 2188 pagina's Aanhangsel 400 pagina's en neem uit elke pagina een fragment van vier zinnen. In deze doelstelling zitten, naast de al in paragraaf 5.1.4 genoemde arbitraire beslissing over de omvang van de steekproef, drie andere arbitraire beslissingen verborgen over de samenstelling van de steekproef.

1. *fragmenten en geen afzonderlijke zinnen.* In het kader van deze studie is niets in te brengen tegen het selecteren van afzonderlijke zinnen. Maar, om de mogelijkheid open te houden dat dit onderzoeksmateriaal ook bestudeerd zou worden op andere verschijnselen, bijvoorbeeld afwisseling in zinslengte in elkaar opvolgende zinnen of pronominale verwijzingen over de zinsgrens heen, zijn - zoveel mogelijk - fragmenten genomen van op elkaar volgende zinnen (zie over dit 'zoveel mogelijk' verder onder 5.2.3).

2. *fragmenten van vier zinnen.* In onderzoeken naar leesbaarheid van teksten wordt vaak gewerkt met teksteenheden van ongeveer 100 woorden of teksteenheden met een aantal zinnen die deze omvang het dichtst benaderen. Zie onder anderen *Van Hauwermeiren* (1975:112). Bij de steek-

proeftrekking voor dit onderzoek is ook de grens van 100 woorden als uitgangspunt genomen om eventueel voortgezet onderzoek naar de leesbaarheid van de geselecteerde fragmenten te vergemakkelijken. Bij een globale verkenning van het gedeelte van de correspondentie tussen regering en Staten-Generaal waaruit de steekproef is getrokken, bleek de gemiddelde zinslengte ongeveer 25 woorden te zijn. Daarom zijn steeds fragmenten van vier zinnen geselecteerd. Dit was gemakkelijker en minder tijdrovend dan bij elk fragment de woorden te tellen en te bepalen met hoeveel zinnen de grens van 100 woorden het dichtst zou worden benaderd. Uiteraard verschilt nu de lengte van de fragmenten, in woorden gemeten, vrij sterk, afhankelijk van de lengte van de vier elkaar opvolgende zinnen. Maar zo werd het in elk geval mogelijk, een aantal fragmenten te krijgen van om en nabij de 100 woorden.

3. *uit elke geselecteerde pagina één fragment.* Deze laatste arbitraire beslissing behoeft nauwelijks toelichting. De doelstelling was, zoveel mogelijk verschillende onderwerpen en soorten kamerstukken in de steekproef te krijgen. En verder zal het, gezien de doelstelling betreffende de omvang van steekproef en fragmenten, duidelijk zijn dat er 400 pagina's moesten worden geselecteerd.

5.2.2 Het selecteren van de pagina's

Gesteld dat voor deze steekproef van 400 pagina's uit ongeveer 20.000 pagina's Bijlagen en 2188 pagina's Aanhangsel elke vijfentwintigste pagina zou worden getrokken, dan valt niet te bewijzen dat de samenstelling van de Bijlagen (bij het Aanhangsel deed zich dit probleem niet voor) niet van invloed is geweest op de uiteindelijke selectie. In de Bijlagen is namelijk een regelmatig patroon te herkennen bij de teksten naar aanleiding van een wetsontwerp: memorie van toelichting, voorlopig verslag, memorie van antwoord enz. Daarom is hier een selectiemethode gevolgd waarbij de samenstelling van het materiaal geen invloed uitoefent, namelijk een aselechte steekproeftrekking. Dit wil zeggen, dat de pagina's zijn geselecteerd aan de hand van een lijst van willekeurige getallen (opgesteld volgens een standaardprogramma 'toevalsgetallen' door het Mathematisch Centrum in Amsterdam).

Bij de selectieprocedure zijn, behalve de pagina's uit registers en inhoudsopgaven, ook de pagina's uit wetsteksten overgeslagen. Deze teksten waren gemakkelijk te herkennen aan de artikelsgewijze opzet, de aanhef en de slotclausule. Tot de wetsteksten zijn gerekend: wetsontwerpen of ontwerpen tot wijziging en de koninklijke boodschap of de begeleidende brief hierbij (de definitieve wetsteksten behoren niet tot de Bijlagen).

Over de omvang van de pagina's die betrekking hebben op wetsteksten, bestonden geen gegevens. Op het tijdstip dat aan de selectieprocedure voor de Bijlagen werd begonnen, was alleen bekend dat het aantal pagina's Bijlagen in het kalenderjaar 1975 20.011 bedroeg en dat het aantal pagina's elk jaar toeneemt. Daarom is voor het nemen van de steekproef uit de Bijlagen, zonder de hierboven vermelde wetsteksten, het aantal pagina's heel ruim geschat op 20.000. Na de selectie van de pagina's verscheen het Jaarverslag over 1976 van het Staatsdrukkerij- en Uitgeverijbedrijf. Hierin staat op pagina 20 het volgende: "In het kalenderjaar 1976 werden 21.191 pagina's kamerstukken gepubliceerd tegenover 20.011 pagina's in het voorafgaande jaar."

Uit de Bijlagen zijn pagina's geselecteerd aan de hand van een lijst van 500 willekeurige getallen tussen 1 en 20.000. Dit zou al 500 pagina's opleveren. Maar het getal 20.000 was heel ruim bepaald, op grond van het vermoeden dat achteraf nog ongeveer 20 procent van de geselecteerde pagina's zou afvallen. Ook moesten namelijk buiten beschouwing blijven: pagina's die voornamelijk uit tabellen en grafieken bestaan en pagina's uit andere categorieën teksten, bijvoorbeeld rapporten van ambtelijke commissies die als bijlagen bij kamerstukken worden opgenomen (zie pagina 90). Omdat de Bijlagen alleen per volgnummer zijn gepagineerd, moesten de ruim 21.000 pagina's met de hand worden geteld. Uiteraard kunnen bij dit handwerk fouten zijn gemaakt, doordat pagina's zijn overgeslagen of dubbel zijn geteld. Maar de eventuele fouten zijn willekeurig en hebben dus geen invloed op de selectieprocedure.

Uit 17.606 pagina's - de overige pagina's maakten dus deel uit van de wetsteksten of inhoudsopgaven e.d. die tijdens de selectieprocedure buiten beschouwing waren gelaten - bleken via deze procedure 439 pagina's te zijn geselecteerd. Dus 1 op de 40 pagina's van de vastgestelde populatie uit de Bijlagen doet mee in de steekproef. Voor de selectie van pagina's uit het Aanhangsel is op soortgelijke wijze te werk gegaan met een verhouding van 1 op 40 aan de hand van een lijst van 55 willekeurige getallen tussen 1 en 2200. Dit resulteerde in 54 pagina's

Uiteindelijk leverden dus de Bijlagen en het Aanhangsel 439 plus 54 is 493 pagina's. Maar van deze pagina's is nog ongeveer een vierde deel niet bij de selectieprocedure voor de fragmenten betrokken.

Van de 439 pagina's die volgens de selectieprocedure uit de Bijlagen waren geselecteerd, zijn namelijk 111 buiten beschouwing gelaten, omdat ze voornamelijk bestonden uit overzichten en tabellen, of omdat ze uit teksten kwamen die het karakter hadden van een wet of een formulier,

bijvoorbeeld een amendement zonder toelichtende tekst, of omdat ze behoorden tot teksten die in eerste instantie niet bedoeld waren als correspondentie tussen regering en Staten-Generaal. Over de laatste categorie het volgende. Het betreft hier pagina's uit teksten die niet door departementen of kamercommissies voor regering of Staten-Generaal werden geschreven. Deze teksten krijgen vaak achteraf - maar het beleid hierin is wisselvallig - de status van kamerstuk, omdat het presidium van de Tweede Kamer besluit een rapport dat aan de Tweede Kamer wordt toegezonden als kamerstuk af te drukken. Deze teksten zijn niet in de steekproef opgenomen, omdat het aanvankelijk in de bedoeling lag deze in wezen andere categorie teksten via een andere steekproef te bestuderen. Bedoeld zijn onder andere de volgende teksten: het jaaradvies van de Raad voor het Wetenschapsbeleid, een rapport van het Economisch Instituut voor de Bouwnijverheid en het Rapport van de Commissie van Drie.

Zo bleven uit de Bijlagen uiteindelijk 439 min 111 is 328 pagina's over. Van de 54 pagina's uit het Aangangsel vielen er ook nog twee af. Het totaal aantal pagina's uit Bijlagen en Aangangsel dat in aanmerking kwam voor het selecteren van de fragmenten kwam zo op 380 (328 plus 52).

Overigens blijkt uit het volgende voorbeeld dat een aselechte steekproef nog niet representatief hoeft te zijn. In de steekproef zitten 16 pagina's moties en 14 pagina's amendementen. Omgerekend over 17.606 pagina's zou dit neerkomen op 643 moties en 561 amendementen. Maar de griffie van de Tweede Kamer deelde mij mee, dat er in het parlementaire jaar 1975-1976 387 moties en 755 amendementen zijn ingediend. Moties en amendementen beslaan bijna altijd één pagina. Bij controle bleek dat de moties 392 pagina's en de amendementen 794 pagina's in beslag namen. De aantallen moties en amendementen in de steekproef zijn dus niet representatief voor de aantallen in de populatie.

5.2.3 Het selecteren van de fragmenten

Uitgangspunt was: uit elke geselecteerde pagina één fragment van vier zinnen. Maar het aantal zinnen per pagina verschilde sterk, van 1 tot 30. Daarom leek het beter, alle geselecteerde pagina's samen als één grote doorlopende pagina te beschouwen en bij de selectie van de zinnen steeds door te tellen. Zo kreeg elke zin uit elke pagina evenveel kans om in de steekproef terecht te komen.

Na enkele proefnemingen op 90 pagina's bleek, dat die pagina's ruim

100 fragmenten opleverden als uit elke reeks van twaalf zinnen één fragment werd geselecteerd. Uit elke reeks van twaalf zijn steeds de eerste vier zinnen genomen. Een voorbeeld. Neem drie geselecteerde pagina's met respectievelijk 3, 20 en 11 zinnen. Deze pagina's worden dan beschouwd als één pagina van 34 zinnen. Van elke reeks van twaalf zinnen zijn de eerste vier opgenomen in de steekproef. Bij het tellen zijn dus steeds vier zinnen opgenomen en acht zinnen overgeslagen. Volgens deze procedure komen er uit de drie geselecteerde pagina's drie fragmenten en moet op de vierde geselecteerde pagina begonnen worden met twee zinnen overslaan. Wanneer een geselecteerde pagina midden in een zin begon of eindigde, zijn deze afgebroken zinnen niet meegeteld, tenzij de zin doorliep op de volgende pagina. (De steekproef bevatte een paar keer twee opeenvolgende pagina's.)

De methode van het 'doortellen' heeft twee nadelen. Ten eerste, de selectieprocedure oefent nu een moeilijk te bepalen invloed uit op de lengte van de geselecteerde zinnen. Immers, een lange zin heeft meer kans om afgebroken te worden dan een kortere zin. Dus, de steekproef zal naar verhouding een iets lagere gemiddelde zinslengte hebben dan de populatie waaruit de steekproef is getrokken. Ten tweede, op deze manier komen er 'fragmenten' die niet bestaan uit vier op elkaar volgende, samenhangende zinnen, maar uit zinnen van verschillende pagina's. Deze nadelen zijn evenwel kleiner dan het nadeel dat ontstaat, wanneer elke pagina met minder dan vier zinnen buiten beschouwing moet blijven of wanneer uit pagina's met 30 zinnen slechts één fragment wordt geselecteerd.

Voor het selecteren van de fragmenten kwam alleen de eigenlijke tekst in aanmerking. Kopjes, voetnoten, rubricerende aanduidingen enz. zijn overgeslagen. Ook zijn buiten beschouwing gebleven de zinnen waarin een opsomming staat die typografisch is aangegeven doordat elk lid van de opsomming op een nieuwe regel begint. Op dit punt is afgeweken van de procedure die de Werkgroep Frequentie-Onderzoek in een enkel geval heeft gevolgd. Uit het Verslag (1974:82,83) blijkt dat de Werkgroep dit soort zinnen wel heeft meegeteld: "... in een enkel geval wordt een puntkomma of komma als zinseinde beschouwd, namelijk in die gevallen dat deze leestekens als grens optreden van een puntsgewijze opsomming waaraan een dubbele punt voorafgaat." En even verder ter toelichting van een voorbeeld: "Soms ontbreken zelfs komma's ter afscheiding van leden van een dergelijke opsomming. Dan wordt de spatie als zinseinde beschouwd." Maar mijns inziens verschilt dit soort zinnen qua opbouw zo sterk van andersoortige zinnen, dat het mij beter leek deze typografisch aangegeven opsommingen binnen één zinsverband buiten be-

schouwing te laten. Deze zinnen hebben ook meer het karakter van een overzicht of inhoudsopgave. Dit nog afgezien van het feit dat ik in dit soort zinnen liever de dubbele punt vóór de opsomming als zinsgrens zou willen nemen. Overigens kwamen dit soort zinnen slechts zelden voor.

De selectie van de fragmenten is niet foutloos verlopen. Wanneer men over een lengte van 380 pagina's steeds de eerste vier zinnen van elke reeks van twaalf zinnen eruit licht, bestaat de kans dat men een tel-fout maakt of een zinseinde over het hoofd ziet. Bij controle na de selectie bleek inderdaad dat er vier fragmenten waren die vijf zinnen bevatten. Uit deze fragmenten is telkens de laatste zin geschrapt. En na de codeerwerkzaamheden en de computerbewerking bleek dat nog enkele fragmenten vijf of drie zinnen bevatten. Omdat de woorden uit de zo verkregen extra zinnen al verwerkt waren in de frequentietabellen, zijn deze zinnen niet geschrapt. Maar de fouten die bij het handwerk zijn ontstaan, hebben geen invloed op het uitgangspunt bij de selectieprocedure, namelijk dat elke zin van elke geselecteerde pagina even veel kans moest hebben om in de steekproef terecht te komen.

Een andere fout had wel gevolgen. Bij de proefneming waarop de verhouding 'vier op twaalf' was gebaseerd, waren de dichtbedrukte driekolomspagina's uit het Aanhangsel ondervertegenwoordigd, zo bleek achteraf. Dus was het getal twaalf veel te laag. Dit had tot gevolg dat de doelstelling 'uit elke geselecteerde pagina één fragment' uiteindelijk niet is gerealiseerd. Achteraf bleek dat uit de 380 pagina's 489 fragmenten waren geselecteerd. In plaats van de bedoelde 40.000 woorden kreeg de steekproef uiteindelijk een omvang van ruim 48.000 woorden.

5.3 Herkomst van het onderzoeksmateriaal met voorbeelden

In dit overzicht is eerst een tweedeling gemaakt tussen fragmenten uit teksten van de regering en die van de Staten-Generaal, opdat ook mogelijke verschillen tussen teksten van departementen enerzijds en verslagen van kamercommissies en vragen van kamerleden anderzijds aan het licht zouden kunnen komen. Voor deze tweedeling waren al in teksten van de regering waarin vragen stonden opgenomen van kamercommissies, de vragen onder 'Staten-Generaal' gerangschikt en de antwoorden of commentaren daarop onder 'regering'.

Van de 489 fragmenten (zie boven) zijn er 95 die zinnen bevatten uit meer dan één geselecteerde pagina, dus mogelijkwerwijs uit verschillende

categorieën teksten. En verder bestonden enkele fragmenten uit drie of vijf zinnen. Daarom staan in dit overzicht de aantallen zinnen per tekstsoort.

Regering	aantal zinnen
- nota's	405
- memories van toelichting	382
- memories van antwoord	232
- antwoorden op vragen	209
- brieven	<u>188</u>
	1416
Staten-Generaal	
- verslagen	445
- vragen	<u>94</u>
	539
	1955
	=====

Deze zinnen beslaan tekstgedeelten van in totaal 48.257 woorden. Van de zinnen uit teksten van de regering bedraagt het aantal woorden 35.300 en van de zinnen uit teksten van de Staten-Generaal 12.957.

Op de volgende drie pagina's staan voorbeelden afgedrukt van correspondentie tussen regering en Staten-Generaal. Het eerste voorbeeld betreft een pagina uit de memorie van antwoord op de Wet op de Winkelsluiting. Hieruit zijn de twee omliggende tekstgedeelten geselecteerd. Het tweede voorbeeld is een kamervraag met antwoord. De vragen vallen onder 'Staten-Generaal'. Uit het antwoord ('regering') komt één fragment. Het derde voorbeeld is een pagina die bij de steekproeftrekking buiten beschouwing is gelaten; het betreft hier een bijlage bij een wetswijziging.

verwijzen naar het adres, d.d. 26 september jl., dat Geleen-promotion aan de bijzondere commissie, belast met de voorbereiding van dit wetsontwerp, deed toekomen. Mede dank zij de stimulerende van een weekmarkt op deze middag wordt volgens het adres in Geleen zowel in food- als in de non-food-sector op zaterdagmiddag 22 % van de weekomzet gehaald, waarbij de buitenlandse kopers een grote rol spelen. Bij sluiting zullen ook de Nederlandse kopers gemakkelijk in België terecht kunnen, zodat gesteld kan worden, dat deze omzet voor een belangrijk deel zonder meer verloren zal gaan. Een dergelijke situatie doet zich in andere nabij de Duitse grens gelegen gemeenten voor.

Aan deze bezwaren kan naar mijn mening bij een landelijke regeling voor de sluiting op zaterdagmiddag niet door een vrijstelling worden tegemoet gekomen. Hoewel aan de leden van de A.R.P.-fractie kan worden geantwoord, dat de vrije zaterdagmiddag niet behoort tot het harmonisatiestreven van de EEG, dient er toch op te worden gewezen, dat een vrijstelling voor de grensstreken met in het bijzonder het doel Duitse kopers te trekken op grote bezwaren van de Duitse overheid en de Europese Commissie kan stuiten. Bovendien zou dit tot een ongelijke situatie binnen Nederland leiden, die mijns inziens een onaanvaardbare discriminatie zou inhouden. Men denke slechts aan de verhouding tussen de grensstreken en daaraan weer grenzende gebieden binnen Nederland.

Dat een koopavond, die overigens uiteraard niet verplicht is voor de ondernemers, een substituut voor de zaterdagmiddag zou kunnen vormen, heb ik eerder in ontkennende zin beantwoord.

Ik veronderstel, dat een door de leden van de V.V.D.-fractie in § 1 van het voorlopig verslag gestelde vraag of burgemeester en wethouders ook over de sluiting op zaterdagmiddag kunnen beslissen op een misverstand berust. Door het Bedrijfschap Horeca wordt in zijn adres niet gesteld dat burgemeester en wethouders deze bevoegdheid zouden hebben. Indien de winkels volgens de voorgestelde regeling op zaterdagmiddag open mogen blijven, hebben burgemeester en wethouders niet de bevoegdheid dit te verhinderen. Bij verplichte sluiting op deze middag zou de gemeenteraad een vrijstelling wegens plaatselijke omstandigheden kunnen geven. Zoals ik reeds aangaf komt het mij niet waarschijnlijk voor dat een dergelijke verordening voor goedkeuring in aanmerking zou kunnen komen, indien deze uitsluitend is gebaseerd op de plaatselijke omstandigheden van ligging nabij de Duitse grens.

Een aantal andere leden van de K.V.P.-fractie informeerden naar de stress-situatie, die volgens de Raad voor het Winkelbedrijf op zaterdag zal ontstaan. Ik acht het bepaald niet uitgesloten dat bij invoering van een sluiting op zaterdagmiddag een zodanige verschuiving van de aankopen naar de zaterdagmorgen ontstaat, dat voor het (eventueel uitgebreide) winkelpersoneel een onaanvaardbare belasting ontstaat. Voorts is het te verwachten dat door de concentratie van zeer veel aankopen in slechts enkele uren de zich nu reeds voordoende problemen bij het parkeren in en bij het vervoer naar de binnensteden worden vergroot. Noch voor de werknemers, noch voor de consumenten is een dergelijke overbelasting gewenst.

Naar aanleiding van de vraag van dezelfde leden omtrent een eventueel te verwachten ontslag van personeel en omzetsdaling, wil ik het volgende opmerken. Het komt mij onvermijdelijk voor, dat een sluiting op zaterdagmiddag, dus gedurende een van de meest produktieve verkooptijdvakken, voor een groot deel van de detailhandel tot ontslag van werknemers zal leiden. Waarschijnlijk zal dit in het bijzonder de part-timers betreffen, waarbij erop kan worden gewezen, dat anders dan wel wordt verondersteld, in de detailhandel werkzame part-timers voor het overgrote deel meer dan 13 uur per week werken. Het gaat hier dus niet om incidentele hulpkrachten, die een kleine bijverdienste verkrijgen. Daar staat tegenover dat een gedeelte van de bestedingen, die oorspronkelijk op zaterdagmiddag plaats hadden, verlegd zal worden naar andere tijden, bij voorbeeld naar de wekelijkse koopavond of naar de zaterdagochtend. Indien de personeelsbezetting op die tijden

Zitting 1975-1976

Aanhangsel van de Handelingen

Vragen gesteld door leden van de Kamer, met de daarop door de Regering gegeven antwoorden

1341

Vragen van de leden Van der Sanden, Gardeniers-Berendsen en Krosse (allen K.V.P.) over de activiteiten van lokale omroepen. (Ingezonden 1 juni 1976)

1
Wilt de Minister aangeven welke activiteiten van de Kabel Omroep Deventer, genoemd in onze vragen ¹ onder 4a t/m c, hetzij als programma-onderdeel, hetzij als 'follow-up' van een uitgezonden programma in de praktijk hebben gefungeerd?

2
Is de Minister van oordeel, dat een lokale omroep als zelfstandig communicatiemedium ook nog behoefte heeft aan een ander medium, nl. een kabelkrant, overigens nog afgezien van de klacht van de voorzitter van de Nederlandse Nieuwsblad Pers, die ernstige bezwaren heeft geuit tegen de uitgifte van een krant, gefinancierd uit omroepgelden?

3
Hoe verklaart de Minister, dat hij, zoals in zijn antwoord op onze vragen 4, 5 en 6 wordt betoogd, 'de lokale omroepexperimenten doelbewust zo veel mogelijk vrijheid heeft gelaten om ze de kans te geven zich in uiteenlopende richtingen te ontwikkelen', terwijl deze experimenten toch plaatsvinden op basis van de Omroepwet?

4
Waarom is het financieren van andere activiteiten dan omroepactiviteiten naar het oordeel van de bewindsman niet strijdig met artikel 58 van de Omroepwet indien het om een experiment gaat? Is de Minister van oordeel,

dat in het algemeen bij experimenten van de wet mag worden afgeweken?

¹ Aanhangsel Handelingen, nr. 1084.

Antwoord

Antwoord van Minister Van Doorn (Cultuur, Recreatie en Maatschappelijk Werk). (Ontvangen 1 juli 1976)

1
De onder a t/m c genoemde categorieën van activiteiten kunnen niet in hun algemeenheid als programma-onderdeel of als follow-up gekenschetst worden. Inspraakavonden en acties bijvoorbeeld zijn nu een onderwerp van een programma geweest, dan weer als gevolg van een programma van de grond gekomen.

De openbare bijeenkomsten voor de bewoners van de Deventer binnenstad, die ten doel hebben ook die mensen die geen kabelaansluiting hebben, te betrekken bij de programma's van de KOD door middel van centraal afgedraalde videoprogramma's, vormen op zich doorgaans weer een onderdeel van volgende video-uitzendingen.

2
Het voert naar mijn mening te ver om — zoals met de vraag wordt gesuggereerd — de uiterst Incidenteel (in 1975 3 keer) verschijnende Kabelkrant als een zelfstandig naast de lokale omroep functionerend communicatiemiddel te beschouwen. Met het uitgeven van de kabelkrant wordt mijns inziens een vorm van verslaggeving van en toelichting op de activiteiten van de kabelomroepvereniging nagestreefd, waar wel degelijk behoefte aan kan bestaan.

Of dit zo is, staat geheel ter beoordeling aan de desbetreffende lokale omroepinstelling.

3
De kabelomroepexperimenten vinden hun basis in artikel 48, vijfde lid, van de Omroepwet. Dit bevat de bepaling dat de Minister van CRM in overeenstemming met de Staatssecretaris van Verkeer en Waterstaat regelen kan stellen. Aangezien genoemd artikel overigens geen beperkende bepalingen bevat, was het mogelijk de lokale omroepexperimenten de nodige vrijheid te laten.

4
Bij de beantwoording van eerdere vragen heb ik het oordeel uitgesproken dat ik het financieren van de in vraag 1 bedoelde activiteiten niet 'onjuist of strijdig met de Omroepwet' acht. Beide kwalificaties waren namelijk gebruikt in de desbetreffende vragen (resp. 5 en 6). Overigens is artikel 58 van de Omroepwet niet van toepassing op de kabelomroepexperimenten, omdat het slechts betrekking heeft op instellingen die (ether-)zendtijd verkregen hebben.

In artikel 2, derde lid, van de beschikking van 24 december 1971, waarin de kabelomroepexperimenten mogelijk worden gemaakt, is bepaald dat het door de Minister van CRM toestaan van het gebruik van draadomroepinrichtingen op zich geen aanspraak doet ontstaan op enige vergoeding van kosten door deze minister. Dat desalniettemin door mijn ambtsvoorganger besloten is een deel van de kosten van de experimenten te subsidiëren uit de reserve regionale omroep betekent niet dat artikel 58 Omroepwet van toepassing zou zijn.

Artikel 8c

De vergoeding, bedoeld in artikel 8b, en de uitgaven ter bestrijding waarvan deze is toegekend, blijven buiten beschouwing bij de toepassing van de artikelen 43, 44, 47 en 73.

Artikel II

In de Lager-onderwijswet 1920 worden de volgende wijzigingen aangebracht:

A

1. De artikelen 3bis en 3ter vervallen.
2. Na artikel 3 worden nieuwe artikelen 3bis, 3ter en 3quater ingevoegd, luidende:

Artikel 3bis

1. Onze minister, de Onderwijsraad gehoord, kan goedkeuren dat van de voorschriften gegeven bij of krachtens deze wet wordt afgeweken ten behoeve van de bijzondere inrichting van het onderwijs op een school. Een verzoek om deze goedkeuring wordt ingediend door burgemeester en wethouders indien het een openbare school betreft, en door het schoolbestuur indien het een bijzondere school betreft. Het verzoek vermeldt het doel en de inhoud van de bijzondere inrichting van het onderwijs.
2. Onze minister kan zijn goedkeuring geven voor een bepaalde tijd en aan zijn goedkeuring voorwaarden verbinden.

Artikel 3ter

1. Onze minister kan aan het gemeentebestuur of aan het bestuur van een bijzondere school een vergoeding uit 's rijks kas toekennen:
 - a. ten behoeve van een school, waarvoor de in artikel 3bis bedoelde goedkeuring is verleend, indien de bijzondere inrichting van het onderwijs op die school noodzaakt tot aanschaffing van aangepaste leermiddelen of het treffen van een andere voorziening;
 - b. ten behoeve van een school waarvoor artikel 3bis niet is toegepast, maar een aanschaffing of voorziening, als bedoeld onder a, noodzakelijk is in verband met deelneming van die school aan een onderwijsvernieuwingactiviteit of een stimuleringsproject.
2. Het verzoek om toekenning van de vergoeding gaat vergezeld van een begroting. Een verzoek ten behoeve van een openbare school wordt ingediend door burgemeester en wethouders.
3. Tegen de beslissing van Onze minister kunnen burgemeester en wethouders onderscheidenlijk het bestuur van de bijzondere school bij Ons beroep instellen. Het beroep moet worden ingediend binnen dertig dagen te rekenen van de dag waarop de beschikking van Onze minister is verzonden.

Artikel 3quater

De vergoeding, bedoeld in artikel 3ter, en de uitgaven ter bestrijding waarvan deze is toegekend, blijven buiten beschouwing bij de toepassing van de artikelen 55, 55bis, 55ter en 101.

B

Artikel 25, vijfde lid, vervalt.

C

Artikel 90, zesde lid, vervalt.

5.4 Tekstbewerking

Na de steekproeftrekking zijn de woorden gecodeerd volgens het codeersysteem zoals dat beschreven staat in hoofdstuk 4. De codeerwerkzaamheden verliepen in vijf fasen.

fase 1. De fragmenten op ponskaarten

Deze fase behoeft alleen toelichting in zoverre er bepaalde instructies zijn gegeven voor het uittikken van de fragmenten. De instructies hadden betrekking op het gebruik van haakjes (hier weergegeven door / en /), op de nummering van de regels, op transcripties en op het gebruik van spaties of het juist niet gebruiken daarvan (oneigenlijke spaties).

a. haakjes

Zet tussen haakjes: elk fragment; elk fragmentnummer; de vier zinnen binnen een fragment; elke zin binnen een fragment. De fragmenten kregen zo de volgende structuur:

```
//nummer/ //zin 1/ /zin 2/ /zin 3/ /zin 4///
```

Door deze haakjes blijven de fragmenten onderscheiden en wordt binnen elk fragment het fragmentnummer onderscheiden van de zinnen. Bovendien zijn op deze manier de vier zinnen per fragment te onderscheiden.

b. regelnummers

Voor correcties en voor het later samenvoegen van woorden en hun codes, was het handig elke ponskaart een regelnummer te geven. Vandaar de volgende instructie:

- Begin elke ponskaart met het fragmentnummer en een regelnummer. De fragmenten kregen zo de volgende structuur: (fragment 111)

```
//11101/ //zin 1  
/11102/ zin 1  
/11103/ zin 1/ /zin 2  
/11104/ zin 2/ /zin 3  
/11105/ zin 3  
/11106/ zin 3  
/11107/ zin 3/ /zin 4  
/11108/ zin 4  
/11109/ zin 4///
```

c. transcripties

Ponskaarten bevatten geen kleine letters. Ook staan enkele diacritische tekens en symbolen die we wel in teksten tegenkomen, niet op een ponskaart. Vandaar de volgende instructies:

- Een hoofdletter in de tekst wordt een asterisk plus die hoofdletter op de ponskaart: H wordt ★H.

- Een letter met een diacritisch teken wordt een uitroepteken plus die letter. Bijvoorbeeld: *coördineren* wordt CO!ORDINEREN; *Tracé* wordt *TRAC!E.
- Gebruik het restteken // voor het uitroepteken, het kwadraatteken, het vermenigvuldigingsteken, het paragraafsymbool enz. Tik voor het teken & 'EN'. En om eventueel verwarring tussen woorden en letters te voorkomen volgde ook nog de instructie om de symbolen voor gulden 'f' en 'Fl' uit te tikken als 'FLO'.

d. spaties

- Tik een spatie voor en na elk woord en voor en na de leestekens: , ; : . (als zinseinde) ? ! haakjes () tenzij in een woord, en aanhalingstekens.
 - Tik geen spatie aan het begin van de zin (tussen het openingshaakje en de hoofdletter) en aan het eind van de zin (tussen de punt en het sluitingshaakje).
 - Er mag geen spatie staan tussen gelijk gerichte haakjes.
- Het nauwkeurig spatiëren is van groot belang omdat in een later stadium de spatie achter een woord de aanduiding is, dat er een code toegevoegd moet worden vanaf een andere ponskaart. Vandaar ook de spaties voor en na leestekens, anders zou het leesteken tussen een woord en zijn code komen te staan.

e. oneigenlijke spaties

Hieronder verstaan we de spaties in eigennamen, getallen en anderstalige woordgroepen. Deze woordgroepen of getallenreeksen worden in het Co-deervoorschrift als één woord beschouwd. Vandaar de volgende instructie:

- Vervang de oneigenlijke spaties door plustekens. Bijvoorbeeld: *Raad van State* wordt *RAAD+VAN+*STATE; *50 000* wordt 50+000; *gentlemen's agreement* wordt GENTLEMEN'S+AGREEMENT; *2 miljoen* wordt 2+MILJOEN.

Tot slot zij nog vermeld dat bij deze tekstbewerking geen rekening is gehouden met afbrekingstekens, cursiveringen en alinea-overgangen. Het afbrekingsteken mag niet verward worden met het weglatingsteken, bijvoorbeeld *in- en uitvoer*. In zo'n geval zijn er haakjes gebruikt, *in(- en uit(voer* om ervoor te zorgen dat in de frequentielijsten uiteindelijk *invoer* en *uitvoer* worden opgenomen (zie fase 5).

fase 2. Controle op fase 1

De op ponskaart gezette fragmenten zijn afgedrukt op kettingformulieren en gecontroleerd aan de hand van de volgende vragen:

Zijn alle fragmenten volledig uitgetikt? Komt de getikte tekst overeen met de oorspronkelijke tekst? Zijn alle instructies opgevolgd? Zijn er typefouten gemaakt?

In deze controlefase bleek dat er in de oorspronkelijke tekst enkele zetfouten stonden: 'de aan woord oord zijnde leden' (de aan het woord zijnde leden) 'praktijd' (praktijk) enz. Dit soort fouten zijn verbeterd. Stijlfouten die het gevolg zijn van een onvoldoende taalbeheersing zijn niet verbeterd. Ook variaties in spelling, 'Oost-Europese' naast 'Oosteuropese', 'materiaal intensief' naast 'materiaalintensief' enz. heb ik laten staan. Alleen de kennelijke zetfouten zijn gecorrigeerd.

Na het aanbrengen van correcties (en een controle daarop) is de verbeterde tekst uitgedraaid met witregels tussen alle regels, opdat de codes boven de woorden konden worden aangebracht.

fase 3. *Het coderen*

Codeerassistenten konden, na een bespreking van het Codeervoorschrift en een inwerkperiode, ongeveer 95 procent van de woorden coderen. De overgebleven woorden, de moeilijke gevallen, heb ik zelf gecodeerd. Twijfelgevallen zijn voorgelegd aan de redacteur van *Woordfrequenties*. Na een uitgebreide controle en het aanbrengen van correcties zijn de codes op ponskaarten getikt.

fase 4. *De codes achter de woorden*

De ponskaarten waarop de fragmenten stonden en de ponskaarten met de codes werden samengevoegd, zodanig dat achter elk woord of achter elke reeks woorden die volgens het Codeervoorschrift als één woord moet worden opgevat, de desbetreffende code werd geplaatst. Ter controle werd de tekst met codes afgedrukt op kettingformulieren. Bij deze controle is nagegaan of er fouten waren gemaakt in het op ponskaart zetten van de codes en of de codes op de juiste plaats stonden. Na het aanbrengen van correcties en controle daarop is de tekst op magneetband gezet voor verdere bewerking.

fase 5. *Lijst gecodeerde woorden*

Voor het bepalen van frequenties is de tekst gesplitst in afzonderlijke woorden met de bijbehorende code. Het afdrukken van een lijst van woorden gerangschikt naar code, met het nummer van de regel waarin het woord voorkwam, gaf nog een laatste mogelijkheid tot correctie.

In deze fase zijn de woorden die bij elkaar behoren, dus de woorden met een code van vier cijfers, aan elkaar gekoppeld (zie pagina 68). Bijvoorbeeld: *Hij 300 voegde 2550 daar 5401 nog 500 iets 400 aan 6106 toe 6205*. Een woord met een code waarvan het vierde cijfer een 0 is, wordt gekoppeld aan een woord met een code waarvan het vierde cijfer een 5 is; evenzo de woorden met de vierde cijfers 1 en 6. In de fre-

quentiellijsten komt dan te staan: *voegde 255 toe 620 en daar 540 aan 610*. Bovendien krijgt een woord of woorddeel dat eindigt op een haakje, achter zich het eerstvolgende woord of woorddeel na een haakje: *school(-001 en 700 cursus(gelden 001 wordt schoolgelden 001 en 700 cursus-gelden 001* (zie pagina 98).

Rest nog de vraag: Hoe nauwkeurig zijn de codeerwerkzaamheden uitgevoerd? Voor het beantwoorden van deze vraag heeft de redacteur van *Woordfrequenties*, P.C. Uit den Boogaart, een steekproef getrokken uit mijn corpus. De steekproef bestond uit tien fragmenten (40 zinnen met in totaal 988 woorden). Na controle bleek dat deze steekproef vijf codeerfouten bevatte: *vreemdelingen 000 (001)*; *afzonderlijk 100 (150)*; twee maal *iets doen toekomen 210 (200)*; *iemand is aangewezen 206 op (216 of 100)*. De juiste code staat tussen haakjes.

Dus 5 op de 1000 woorden zijn fout gecodeerd, met naar verhouding veel fouten bij de werkwoordsvormen. Dit komt overeen met wat in *Woordfrequenties* (pag. 27) naar aanleiding van de steekproefsgewijze controle is opgemerkt: een foutenmarge van ongeveer een half procent.

Na het coderen zijn van de woorden frequentielijsten gemaakt op precies dezelfde manier zoals dat is gebeurd in *Woordfrequenties*.

Van de frequentielijsten is, als Bijlage 2, in deze publikatie alleen opgenomen lijst D₁: woordklassen geordend naar lexicale code, geteld naar tokens (zie paragraaf 4.4). Deze lijst is verreweg het belangrijkste voor het toetsen van de hypothesen genoemd in het volgende hoofdstuk; de andere lijsten zijn verkrijgbaar via het Rekencentrum van de TH-Eindhoven.

Daarna is de steekproef op de Computerafdeling van de Letterenfaculteit van de Universiteit van Amsterdam opgenomen in het systeem *Oog op Tekst*, zoals dat beschreven staat in *Brandt Corstius* (1978:70-75). Dit systeem stelt de onderzoeker in staat om uit een corpus zinnen op te vragen die aan een bepaalde voorwaarde voldoen. Een voorwaarde kan geformuleerd worden als een opeenvolging van woorden of codes, of beide, verbonden door de logische voegwoorden *en*, *of* en *niet*. Bovendien mogen drie variabelen optreden: * een woord, c een code, en - een rij woorden met codes (eventueel leeg). Verder is het bijvoorbeeld mogelijk om aan te geven of zo'n voorwaarde aan het begin van de zin (ff) moet voorkomen. Ter illustratie drie voorbeelden van vragen die met behulp van dit systeem gesteld kunnen worden.

1. Uit de D₁-lijst in *Woordfrequenties* valt u op dat de code 237 alleen in het corpus 'dagbladen' voorkomt, en dan nog met frequentie 1. Een verbogen voltooid deelwoord van een hulp- of koppelwerkwoord, kan dat?

Welk woord zou dat kunnen zijn? U vraagt alle zinnen op, die voldoen aan de voorwaarde $\ast 237$. U krijgt dan één zin waarin de zinswending "de ... achterop geraakte 237 New Ulster Movement" voorkomt.

2. Gesteld, u wilt weten of het woord 'aanzien' ook in andere omgevingen voorkomt dan in de combinatie 'ten aanzien van'. U moet dan de zinnen opvragen, die voldoen aan de voorwaarde *aanzien 000 en niet ten c aanzien c van c*.

3. U wilt weten of er in het corpus 'dagbladen' meer zinnen met een persoonsvorm als tweede woord voorkomen dan in het corpus 'populair-wetenschappelijke boeken'. Een persoonsvorm kan als code krijgen:

24., 25., 26. of 27. Het antwoord op deze vraag krijgt u met behulp van de volgende formulering:

$\ast\ast \ast c \ast 24$ of $\ast\ast \ast c \ast 25$ of $\ast\ast \ast c \ast 26$ of $\ast\ast \ast c \ast 27$. Het teken $\ast\ast$ staat voor 'begin van de zin'; $\ast c$ wil hier zeggen 'het eerste woord (met code)'; het tweede teken \ast in elke reeks geeft, door de code erachter, de persoonsvorm aan.

Tot slot zijn, eveneens op de Computerafdeling van de Letterenfaculteit van de Universiteit van Amsterdam, retrogradelijsten met frequenties geproduceerd. In een retrogradelijst staan woorden alfabetisch gerangschikt op de eindletter(s). Verder zijn overzichten gemaakt van zinslengte (het aantal codes per zin) en woordlengte (het aantal letters per woord).

Met behulp van deze gegevens en dank zij de mogelijkheden die *Oog op Tekst* biedt, in combinatie met *Woordfrequenties*, konden de hypothesen worden getoetst die vermeld staan in het volgende hoofdstuk.

6.1 Inleiding

In dit hoofdstuk staat beschreven hoe, naar aanleiding van de oordelen over het taalgebruik in teksten van de overheid (hoofdstuk 3), binnen het kader van het in deze studie gekozen codeersysteem (hoofdstuk 4) hypotheses zijn opgesteld over verschillen tussen de correspondentie tussen regering en Staten-Generaal (hoofdstuk 5) en de corpora 'dagbladen', 'opiniebladen' en 'populair-wetenschappelijke boeken' uit *Woordfrequenties*. Zie paragraaf 6.2. In 6.3 vindt u, na een korte uiteenzetting over de wijze waarop de hypotheses zijn getoetst, een samenvattend overzicht van de toetsingsresultaten. In deze inleiding wordt eerst nog gesproken over de oordelen uit hoofdstuk 3 naar aanleiding waarvan binnen het gekozen codeersysteem geen hypotheses kunnen worden opgesteld.

6.1.1 Oordelen buiten het C_3C -kader

Een aantal oordelen zijn niet getoetst omdat het, binnen het kader van het codeersysteem C_3C , niet mogelijk is er hypotheses over op te stellen. Bedoeld zijn:

A. oordelen waarin gebruik wordt gemaakt van vage termen, die niet geherformuleerd kunnen worden in C_3C -terminologie, bijvoorbeeld de oordelen over onhandige ontkenningen, moeilijke woorden, verbredingen, archaïsch en clichématig taalgebruik enz. (zie verder de samenvatting van de algemene oordelen, paragraaf 3.3.2);

B. oordelen waarin gebruik gemaakt wordt van voornamelijk syntactische termen of onderscheidingen, die niet in het lexicale C_3C -systeem zijn opgenomen, bijvoorbeeld: bijzin, tussenzin, tangconstructie, elliptische constructie en herhaling.

Toch is het niet geheel waar, wanneer men zou stellen dat het codeersysteem C_3C niet toereikend is om verschijnselen op het spoor te komen die met deze oordelen te maken hebben. Uit de oordelen van het type A kunnen wel hypotheses binnen het kader van het C_3C -systeem worden afgeleid. Naar aanleiding van het oordeel over archaïsch taalgebruik

kan bijvoorbeeld een hypothese worden opgesteld over het voorkomen van substantieven in de tweede naamval.

Anders ligt het bij de oordelen onder B. Een meer syntactisch gericht codeersysteem zou hier wel resultaat opleveren. Toch kunnen de hier genoemde verschijnselen voor een gedeelte onderzocht worden, en wel via afgeleide oordelen die passen binnen C_3C . Een voorbeeld: ingewikkelde zinsstructuur en zinslengte correleren vaak positief. En zinslengte kan gemakkelijk onderzocht worden als de woorden gecodeerd zijn. Men heeft dan de computer alleen maar de codes te laten tellen tussen zinsbegintekens en zinseindetekens. Ook kan met aanvullend handwerk nog veel onderzocht worden. De lijdende vorm bijvoorbeeld valt te onderzoeken door alle zinnen met vormen van *worden* en *zijn* op te vragen. Dan kan vrij snel bekeken worden welke vormen van deze werkwoorden eigenlijk gecodeerd hadden kunnen worden als hulpwerkwoorden van de lijdende vorm. Zie hierover verder bij de desbetreffende hypothese in paragraaf 6.2.

Overigens was, strikt genomen, het coderen volgens het C_3C -systeem niet nodig geweest voor oordelen waarin bepaalde woorden worden genoemd: *men*, *ik*, *hetgeen* enz. Immers ook zonder een codeersysteem kan een tekst met behulp van een computer worden verwerkt tot een alfabetische woordenlijst met frequenties. Maar bij oordelen over bepaalde woorden gaat het niet altijd om woorden die geen andere betekenis of functie kunnen hebben. Wanneer wij bijvoorbeeld *dat* als relativum willen onderzoeken, is er een codeersysteem nodig om onderscheid te kunnen maken tussen *dat* als relativum, demonstrativum en onderschikkend voegwoord.

6.2 Hypotheses

In deze paragraaf worden, voor zover mogelijk, de oordelen uit hoofdstuk 3 geherformuleerd tot hypotheses waarin gebruik gemaakt wordt van C_3C -terminologie. Bij het opstellen van de hypotheses is de onderverdeling van de oordelen aangehouden, zoals gegeven in paragraaf 3.4, met de trefwoorden: onpersoonlijk, ingewikkeld, wijdlopig en traditioneel. Bij de hypotheses en de bespreking ervan zijn de volgende afkortingen gebruikt:

- TO - teksten van de overheid, in deze studie dus de correspondentie tussen regering en Staten-Generaal;
- R - teksten van de regering uit TO;
- StG - teksten van de Staten-Generaal uit TO;
- D - dagbladen;
- O - opiniebladen;
- W - populair-wetenschappelijke boeken.

De subcorpora R en StG zijn samen als corpus T0 vergeleken met de corpora D, O en W. De toevoeging 'corpus' of 'subcorpus' wordt gebruikt om in twijfelgevallen duidelijk aan te geven dat de steekproef bedoeld is en niet de populatie waaruit de steekproef is getrokken. (Deze aanduiding is weggelaten in tabellen, waar verwarring uitgesloten is omdat getallen altijd betrekking hebben op de steekproef.) Een punt in of achter een code wil zeggen, dat op die plaats elk mogelijk cijfer uit het C₃C-systeem mag staan. Een voorbeeld: met code 00. ('gewone' substantieven) is bedoeld: 000 (basisvorm), 001 (meervoudsvorm), 002 (genitief) en 003 (overige verbogen vormen). Eventuele beperkingen volgen bij de desbetreffende hypothese.

Per hypothese worden de gevonden frequenties van het genoemde verschijnsel vermeld. Voor gedetailleerde informatie over frequenties per code zij verwezen naar Bijlage 2 (pagina 173 tot 176). Bij het vergelijken van de frequenties moet wel bedacht worden dat de omvang van de subcorpora R en StG aanzienlijk kleiner is dan die van de corpora D, O of W. De omvang van de corpora in aantallen woorden is:

D	O	W	T0	(R)	(StG)
120313	123873	119979	48257	35300	12957

6.2.1 Onpersoonlijk

6.2.1.1 Naamwoordstijl

Naar aanleiding van de oordelen onder de noemer 'naamwoordstijl' kan niet zonder meer de hypothese worden opgesteld dat in T0 meer substantieven voorkomen dan in D, O of W. De categorie substantieven in C₃C bevat immers ook eigennamen (01.), interjectivisch gebruikte substantieven (08.) en woorden in zelfnoemfunctie (09.). De oordelen over de naamwoordstijl kunnen binnen de categorie substantieven alleen geherformuleerd worden in een hypothese over de 'gewone' substantieven (00.) en de 'adjectivisch gebruikte substantieven (02.).

Bij deze hypothese zijn de codes met als derde cijfer 2 (genitief) en 3 (overige verbogen vormen) buiten beschouwing gelaten, omdat het oordeel over naamwoordstijl hier geen betrekking op heeft (zie pagina 34). Deze codes komen nog wel aan de orde bij een hypothese naar aanleiding van het oordeel 'traditioneel taalgebruik' in paragraaf 6.2.4. Dus alleen de codes 000, 001, 020 en 021 blijven over. Wanneer we de frequenties van

deze codes optellen (zie Bijlage 2) krijgen we de volgende getallen:

D	O	W	TO	(R)	(StG)
24299	23808	24900	11206	8273	2933

Deze getallen zijn gebruikt bij het toetsen van de volgende hypothese:

1. substantieven

In TO komen de codes 000, 001, 020 en 021 meer voor dan in D, O of W.

Naast deze hypothese binnen de categorie substantieven zijn er nog drie op te stellen over codes binnen andere categorieën die toch gerekend kunnen worden tot de naamwoorden. Ten eerste code 12. voor substantivisch gebruikte adjectieven in welke vorm dan ook; dus alle mogelijke derde cijfers, in dit geval 0, 1, 3, 6, 7 en 9. Ten tweede, code 2.1 voor substantivisch gebruikte infinitieven. Ten derde, de codes die gebruikt worden voor substantivisch gebruikte deelwoorden. Vandaar de volgende drie hypothesen (per hypothese worden de getallen vermeld die gebruikt zijn bij de toetsing):

2. substantivisch gebruikte adjectieven

In TO komen de codes 120, 121, 123, 126, 127 en 129 meer voor dan in D, O of W.

D	O	W	TO	(R)	(StG)
234	259	402	64	47	17

3. substantivisch gebruikte infinitieven

In TO komen de codes 201, 211, 221 en 231 meer voor dan in D, O of W.

D	O	W	TO	(R)	(StG)
242	305	410	211	166	45

4. substantivisch gebruikte deelwoorden

In TO komen de codes 204, 214, 224, 234, 208, 218 en de substantivisch gebruikte deelwoorden in verbogen vorm, meer voor dan in D, O of W.

D	O	W	TO	(R)	(StG)
59	75	68	36	31	5

Bij het opstellen van hypothese 4 deed zich nog een probleem voor. Onder deze deelwoorden vallen niet alleen de woorden met codes 2.4 en 2.8 voor het onvoltooid en voltooid deelwoord in het meervoud, bijvoorbeeld *de stervenden 204* en *de gevallen 208* (de codes 228 en 238 kwamen niet voor), maar ook een gedeelte van de woorden met de codes 2.3 en 2.7 voor het onvoltooid en voltooid deelwoord in verbogen vorm: wel *het volgende 203* maar niet *de volgende 203 opmerkingen*; wel *de verdachte 217* maar niet *een verdachte 217 opmerking*. Daarom moest voor het opstellen van deze hypothese eerst nagegaan worden hoe vaak woorden met codes 2.3 en 2.7 substantivisch gebruikt voorkomen, dat wil zeggen, zonder daarbijbehorend zelfstandig naamwoord.

Via *Oog op Tekst* zijn opgevraagd de zinnen met de codes 2.3 en 2.7 zonder een direct daaropvolgend woord met code 0.; daarna is op basis van het zinsverband nagegaan, of er werkelijk sprake was van substantivisch gebruik. Want de computer gaf op deze vraag ook zinnen met bijvoorbeeld *werkende 203 jongeren 126* of *de genoemde 217 sociale 103 encyclieken 001*. Na het schrappen van deze gevallen bleef er een, beperkt aantal substantivisch gebruikte deelwoorden in verbogen vorm over. In schema:

	D	O	W	TO	(R)	(StG)
1.	822	758	899	597	468	129
2.	114	110	143	78	62	16
3.	16	29	29	14	13	1

1. codes 2.3 en 2.7

2. codes 2.3 en 2.7 zonder direct volgend woord met code 0..

3. substantivisch gebruikte deelwoorden in verbogen vorm

De getallen achter 3 vormen dus samen met de frequenties van deelwoorden in het meervoud (codes 2.4 en 2.8) het materiaal voor het toetsen van hypothese 4 over substantivisch gebruikte deelwoorden.

Onder naamwoordstijl werd ook het oordeel vermeld dat deze stijl gepaard gaat met een armoede aan werkwoorden. Blijkens de opmerkingen op pagina 34 zijn hiermee niet bedoeld: hulp- of koppelwerkwoorden, bijwoordelijk gebruikte deelwoorden en de hierboven genoemde substantivisch gebruikte infinitieven, deelwoorden in het meervoud en deelwoorden in verbogen vorm (al dan niet met bijbehorend substantief). En verder zijn in de hypothese ook buiten beschouwing gelaten interjectivisch gebruikte werkwoorden, imperatieven en 'overige vormen' (voornamelijk de aanvoegende wijs); deze worden nog apart onderzocht via de hypothesen 17, 18 en 39. Deze beperkingen op de categorie werkwoorden resulteerden in de volgende hypothese. Onder de hypothese staan de getallen die gebruikt zijn bij de toetsing.

5. werkwoorden

In T0 komen de codes 2.. met tweede cijfer 0, 1 of 2 (en daarna alleen ..0, ..2 of ..6) en met tweede cijfer 4, 5 of 6 (en daarna geen ..7, ..8 of ..9) minder voor dan in D, O of W.

D	O	W	T0	(R)	(StG)
12329	13051	12086	4660	3318	1342

Verder wordt in de naamwoordstijl volgens de oordelen meer gebruik gemaakt van voorzetsels. Dit verschijnsel kan trouwens ook in verband gebracht worden met tangconstructies. Zie hierover verder onder 6.2.2.2. Hier zijn alleen de codes 600 en 603 genomen. Andere codeermogelijkheden onder 6.. komen nog aan de orde bij de hypothesen 24, 27, 28 en 29.

6. voorzetsels

In T0 komen de codes 600 en 603 meer voor dan in D, O of W.

D	O	W	T0	(R)	(StG)
15732	14676	14816	8039	6063	1976

Het laatste verschijnsel dat onder naamwoordstijl werd behandeld, was het veelvuldig voorkomen van woorden op '-ing'. Voor het opstellen van een hypothese hierover kon gebruik worden gemaakt van een retrograde-programma (zie pagina 101) waarmee alle woorden op '-ing' of '-ingen' konden worden opgevraagd. De retrogradelijsten bevatten ook woorden op '-ing' of '-ingen' met code 1.. (onderling, plotseling) en code 2.. (ging, ontvingen). Deze codes zijn uiteraard niet meegeteld. Ook de eigennamen, code 01., zijn buiten beschouwing gelaten (Groningen). Voorts zijn overgeslagen: buitenlandse woorden (planning, holding) en woorden waarin '-ing' of '-ingen' niet als achtervoegsel kan worden opgevat (ding, kringen). Deze telling resulteerde in de volgende getallen:

D	O	W	TO	(R)	(StG)
2552	2332	2678	2386	1809	577

Deze getallen werden gebruikt voor het toetsen van de volgende hypothese:

7. woorden op '-ing' of '-ingen'

In TO komen woorden op '-ing' of '-ingen' meer voor dan in D, O of W.

Bij het toetsen van deze hypothese (zie paragraaf 6.3) zijn de desbetreffende woorden niet alleen vergeleken met het totaal, maar ook met het aantal woorden van de woordsoort waartoe ze behoren (de codes 000 en 001). Immers als de corpora verschillen in frequenties van 000 en 001, kunnen ze alleen daarom al verschillen in woorden op '-ing' of '-ingen'.

6.2.1.2 Lijdende vorm

De opmerkingen uit hoofdstuk 3 over het veelvuldig voorkomen van de lijdende vorm in teksten van de overheid, kunnen niet direct geherformuleerd worden tot een hypothese. In het codeersysteem C₃C bestaat namelijk geen aparte code voor de hulpwerkwoorden van de lijdende vorm 'worden' en 'zijn'. De vormen van deze werkwoorden krijgen net als die van andere hulp- en koppelwerkwoorden de codes 23. of 27., bijvoorbeeld *hij wordt* 273 *burgemeester* en *hij is* 273 *aanwezig geweest* 236.

Wel is het mogelijk via *Oog op Tekst* de zinnen op te vragen waarin persoonsvormen van 'worden' en 'zijn' als hulp- of koppelwerkwoord

voorkomen. Voor het opsporen van lijdende vormen moeten dan ook opgevraagd worden de zinnen met 'worden' en 'zijn' als infinitief, code 230, bijvoorbeeld *hij kan niet ontslagen worden* 230. Dan kan worden nagegaan, hoe vaak 'worden' en 'zijn' als koppelwerkwoord voorkomen en hoe vaak 'zijn' een hulpwerkwoord van tijd is. De vormen die overblijven zijn dan hulpwerkwoorden van de lijdende vorm.

Toch levert dit geen zinvolle beperking van het aantal te onderzoeken zinnen op. In TO, dat uit 1955 zinnen bestaat, is de frequentie van 'worden' en 'zijn' als hulp- of koppelwerkwoord 1855. (Niet meegerekend zijn de imperatieven en overige vormen: derde cijfer 7, 8 of 9 bij persoonsvormen.) Hieronder volgen ter vergelijking de aantallen uit D, O en W:

	D	O	W	TO	(R)	(StG)
1.	6631	6812	5752	1955	1416	539
2.	3819	4083	4055	1855	1372	483

1. het aantal zinnen per corpus

2. 'worden' en 'zijn' als hulp- of koppelwerkwoord (code 27. met derde cijfer 1 tot en met 6, en code 230)

Dit zou betekenen dat zo het grootste deel van de zinnen uit een corpus weer terugkomt. Uiteraard kan een zin meer dan één lijdende vorm bevatten. Deze zin wordt dan zo vaak afgedrukt als er verschillende vormen van 'worden' en 'zijn' in staan. Betreft het dezelfde vorm, dan wordt de zin maar één keer gegeven. Dus bij elke zin moet nog worden nagegaan of er meer dan één lijdende vorm in zit.

Het zoeken naar andere beperkingen op dit aantal leverde geen resultaat op. Wanneer de eis wordt gesteld dat er in dezelfde zin een voltooid deelwoord moet voorkomen, geeft de computer ook zinnen zonder een lijdende vorm: *Ik ben* 271 *pas ziek geweest* 236 en *Hij wordt* 273, *voor zover ik heb begrepen* 206, *geen burgemeester*. De beperking tot transitieve deelwoorden sluit zinnen uit als *Er wordt* 273 *gedacht* 206 *aan* ... De beperking tot zinnen waarin de 'bepaling met door' voorkomt is evenmin zinvol; de lijdende vorm komt ook voor zonder deze bepaling.

Het lijkt aannemelijk dat er tussen D, O, W en TO geen verschillen bestaan in het voorkomen van vormen van 'worden' en 'zijn' als koppelwerkwoord en van 'zijn' als hulpwerkwoord van tijd. Een hypothese over het voorkomen van de lijdende vorm kan dan worden getoetst op basis van

de op de vorige pagina gegeven getallen. Eventuele verschillen kunnen dan worden toegeschreven aan het voorkomen van de lijdende vorm.

Toch bleef het een uitdaging om na te gaan of het verschil nu werkelijk zit in 'worden' en 'zijn' als hulpwerkwoord van de lijdende vorm. Voor dit doel zijn de lijdende vormen met de hand geteld. Hierbij is uit de corpora D, O en W, een even groot aantal zinnen genomen als TO bevat, namelijk 1955. Uit D, O en W zijn steeds de eerste 1955 zinnen genomen. Bij het tellen van de lijdende vormen is het volgende criterium gehanteerd: De woorden 'worden' en 'zijn' moeten aanwezig zijn met als code 271, 272, 273, 274, 275, 276, en 230 voor de infinitieven, en er moet een voltooid deelwoord aanwezig zijn met als code 206 of 216.

Dit criterium houdt in dat niet worden meegeteld:

1. Lijdende vormen in beknopte bijzinnen. In bijvoorbeeld *een gebied, bestemd 216 voor recreatie en de vraag, in het vorige hoofdstuk gesteld 216* is geen persoonsvorm aanwezig. Hetzelfde geldt voor *de door de minister ontvangen 216 nota*; zie hierover hypothese 31 (klemconstructies).
 2. De constructie met een persoonsvorm van 'zijn' gevolgd door te+infinitief. In bijvoorbeeld *A is 273 in overeenstemming te brengen 210 met B* is geen code 216 aanwezig. (Zie hierover nog de Discussie, pagina 155.)
 3. De aanvoegende wijs. Bijvoorbeeld *Voorts zij 279 opgemerkt*. Deze vorm komt nog aan de orde bij hypothese 39 onder archaïsmen.
- In de enkele gevallen dat er een 'samengestelde lijdende vorm' voorkwam, bijvoorbeeld *Hierbij kon toestemming gegeven 216 en verkregen 216 worden 230*, is deze als één lijdende vorm geteld. Criterium bleef dat er steeds sprake moest zijn van een combinatie van een voltooid deelwoord (206 of 216) en 'worden' of 'zijn' als hulp- of koppelwerkwoord.

Bij het tellen van de lijdende vorm deden zich soms twijfelgevallen voor. Immers 'worden' en 'zijn' kunnen ook koppelwerkwoord zijn en bovendien kan 'zijn' nog voorkomen als hulpwerkwoord van tijd. Zie onder anderen *Pollmann* (1975:23-25, 146-151) en *Kraak en Klooster* (1968:136). Het onderscheid tussen hulpwerkwoord van tijd en hulpwerkwoord van de lijdende vorm gaf weinig problemen. Bij twijfelgevallen is het criterium gehanteerd dat er pas sprake is van een lijdende vorm, wanneer een actief subject in de vorm van een 'bepaling met door' (of een equivalent daarvan, bijvoorbeeld 'onzerzijds' in plaats van 'door ons') aan-

wijsbaar of toelaatbaar is. Vandaar geen lijdende vorm (-) in 1 maar wel (+) in 2.

- 1. De planninggroep is al aan de uitwerking begonnen.
- + 2. Met de bouw is alvast begonnen.

Moeilijker is het onderscheid tussen 'zijn' als koppelwerkwoord en 'zijn' als hulpwerkwoord van de lijdende vorm. Bij twijfelgevallen is steeds gekozen voor koppelwerkwoord. 'Zijn' is pas opgevat als hulpwerkwoord van de lijdende vorm, wanneer van de desbetreffende zin een actieve pendant mogelijk is en wanneer een 'bepaling met door' aanwijsbaar of toelaatbaar is. Dus geen lijdende vorm in de volgende zinnen:

- 3. Hij was betrokken bij een ongeluk.
- 4. Anderen zullen hiermee gediend zijn.
- 5. Deze maatregelen zijn gericht op behoud van het milieu.

Een aparte moeilijkheid vormden de gevallen waarin het woord dat als voltooid deelwoord gecodeerd is, verschillende betekenissen kan hebben. In deze gevallen kon pas na interpretatie van de zin, of soms de voorafgaande en/of volgende zin, een beslissing worden genomen. Enkele voorbeelden:

- 6a. Deze industrieën zijn voor hun afzet aangewezen op het buitenland.
- + 6b. Zij zijn aangewezen als vertegenwoordigers.
- 7a. Zijn taak is beperkt. (Hij heeft een kleine taak.)
- + 7b. Zijn taak is beperkt. (... door zijn chef.)
- 8a. Hij is ervan overtuigd. (Hij is van mening.)
- + 8b. Hij is ervan overtuigd. (... door zijn vrienden.)
- 9a. Dit is uitgewerkt. (Dit tabletje is uitgewerkt.)
- + 9b. Dit is uitgewerkt. (Dit voorstel is uitgewerkt.)

Volgens de hier gegeven criteria en na interpretatie van de context is per corpus ongeveer 20 procent van de constructies met 'zijn' plus voltooid deelwoord beschouwd als constructie met koppelwerkwoord. Telling van 'worden' en 'zijn' als hulpwerkwoord van de lijdende vorm gaf als resultaat:

D	O	W	TO	(R)	(StG)
392	384	492	1128	869	259

Deze getallen werden gebruikt voor het toetsen van de volgende hypothese:

8. Lijdende vorm

In TO komt de lijdende vorm meer voor dan in D, O of W.

6.2.1.3 Het menselijk element

Naar aanleiding van de opmerkingen hierover in hoofdstuk 3 zijn tien hypotheses opgesteld, de nummers 9 tot en met 18.

Wanneer er sprake is van een onpersoonlijke weergave, zullen er minder personalia gebruikt worden. Ook het minder voorkomen van possessiva kan aanleiding geven tot het oordeel 'onpersoonlijke weergave'. Maar voor beide categorieën geldt dat er woorden toe behoren die (kunnen) verwijzen naar zaken in plaats van personen. Bij de personalia zijn dit de woorden *het*, *zij* en *ze*; voor de possessiva de woorden *zijn*, *haar* en *hun*. Bij het woord *het* is aangenomen dat er altijd sprake is van een onpersoonlijke verwijzing omdat, wanneer dit woord slaat op een persoon (bijvoorbeeld 'het kind'), de verwijzing onpersoonlijker is dan via het woord *zij* of *haar*.

Via *Oog op Tekst* zijn de zinnen opgevraagd waarin *zij*, *ze*, *zijn*, *haar* of *hun* voorkomen, opdat bepaald kon worden of deze woorden verwijzen naar zaken of personen. In de volgende tabel is aangegeven hoe vaak deze woorden verwijzen naar zaken. Deze aantallen plus de frequenties van *het* als personale zijn afgetrokken van de frequenties voor personalia of possessiva:

	D	O	W	TO	(R)	(StG)
totaal personalia	2476	3378	2836	319	197	122
het 300/303	139	107	139	16	12	4
ze 300/303	275	384	275	11	3	8
ze (zaken)	60	100	100	11	3	8
zij 300	205	222	385	104	64	40
zij (zaken)	26	47	163	58	45	13
personalia zaken	225	254	402	85	60	25
personalia personen	2251	3124	2434	234	137	97

	D	O	W	TO	(R)	(StG)
totaal possessiva	1110	1373	1293	200	133	67
haar 330	123	105	145	44	27	17
haar (zaken)	42	48	68	42	26	16
hun 330	220	281	328	59	37	22
hun (zaken)	46	41	98	25	19	6
zijn 330	550	608	527	43	26	17
zijn (zaken)	91	72	108	17	11	6
possessiva zaken	179	161	274	84	56	28
possessiva personen	931	1212	1019	116	77	39

De getallen achter 'personalia personen' en 'possessiva personen' werden gebruikt voor het toetsen van de volgende hypothesen:

9. personalia

In TO komen personalia die verwijzen naar personen, een gedeelte van de codes 300, 302 en 303, minder voor dan in D, O of W.

10. possessiva

In TO komen possessiva die verwijzen naar personen, een gedeelte van de codes 321, 323, 330, 332 en 333, minder voor dan in D, O of W.

Maar vooral het ontbreken van verwijzingen naar een 'eerste persoon' kan een tekst onpersoonlijk maken. Vandaar de volgende hypothese:

11. verwijzingen naar een 'eerste persoon'

In TO komen verwijzingen naar een 'eerste persoon' minder voor dan in D, O of W.

Voor het toetsen van deze hypothese zijn de frequenties opgeteld van de volgende woorden: *ik*, *'k*, *mij*, *me*, *mijszelf*, *mezelf*, *mijn*, *m'n*, *wij*, *we*, *ons*, *onze*, *onsszelf*, en alle verbogen vormen (*mijne*, *mijns*, *onzer* enz.). De frequenties van deze woorden zijn

D	O	W	TO	(R)	(StG)
1051	1736	1382	124	95	29

De woorden die verwijzen naar een eerste persoon zijn niet alleen vergeleken met het totaal, maar ook met het aantal woorden van de woordsoorten waartoe ze behoren, namelijk personalia, possessiva en reflexiva.

Het oordeel dat in plaats van *ik* vaak *ondergetekende* wordt gebruikt, leidt tot de hypothese:

12. 'ondergetekende(n)'

In TO komt het woord 'ondergetekende(n)' meer voor dan in D, O of W.

De oordelen over het vaak voorkomen van *men* en over het schaarse gebruik van *u* kunnen gemakkelijk geherformuleerd worden tot hypotheses:

13. 'men'

In TO komt het woord 'men' meer voor dan in D, O of W.

14. 'u'

In TO komt het woord 'u' minder voor dan in D, O of W.

De frequenties van *men* en *u* zijn:

	D	O	W	TO	(R)	(StG)
men	182	317	448	36	22	14
u	73	132	64	33	11	22

Bij het toetsen van deze hypotheses moet, net als bij hypothese 11, rekening gehouden worden met het feit dat de woordsoorten waartoe deze woorden behoren, in frequentie kunnen verschillen. Vandaar dat het voorkomen van *men* ook gerelateerd is aan code 440 (zelfstandige indefinita) en het voorkomen van *u* aan de codes 30. en 34. (personalia en reflexiva).

Onder 'het menselijk element' is op pagina 36 ook het oordeel vermeld, dat in teksten van de overheid weinig zinnen in de vragende vorm voorkomen. Dit oordeel kan niet direct geherformuleerd worden tot een hypothese. Zinnen die eindigen met een vraagteken zijn niet als zodanig gecodeerd. Hetzelfde geldt trouwens voor zinnen met een uitroepeten. Maar gelukkig vertonen vraagzinnen kenmerken, op grond waarvan via *Oog* op *Tekst* wel een hypothese kan worden opgesteld. Meestal beginnen vraagzinnen met een persoonsvorm, of een vraagwoord als eerste, tweede of derde woord. Vandaar de volgende hypothese:

15. vraagzinnen

In TO komen vraagzinnen minder voor dan in D, O of W.

De manier waarop het materiaal is verzameld voor het toetsen van deze hypothese, vergt een uitgebreide toelichting.

Via *Oog op Tekst* zijn eerst alle zinnen opgevraagd die beginnen met een persoonsvorm. Maar van deze zinnen bleek slechts een beperkt aantal echte vraagzinnen te zijn. De overige zinnen betroffen: voorwaardelijke zinnen, *Wil 273 hij veranderen dan moet er heel wat gebeuren*; imperatieven, *Zie 257 pagina 100*; aansporingen, *Laat 271 ik het nog eens uitleggen*; uitroepen, *Kom 247 daar nou eens om!*; toegevingen, *Waren 246 er vorig jaar twee gevallen per week, nu zijn er drie per dag*; en elliptische zinnen, *Belde 245 daarna om zich aan te geven*. In corpus W overtrof het aantal voorwaardelijke zinnen zelfs het aantal vraagzinnen. Verder zij er nog op gewezen dat, door het opvragen van zinnen die beginnen met een persoonsvorm, de ingebedde vraag die niet aan het begin van de zin staat, buiten beschouwing blijft, bijvoorbeeld *Nu er een kans komt om daar medewerkster te worden, zou 275 men dan niet gaan?* De ingebedde vraag aan het begin van de zin kon wel in de tellingen worden betrokken: *Is 273 dit discriminatie? vroeg ik mij af*. Hieronder de gevonden frequenties in schema.

	D	O	W	TO	(R)	(StG)
1.	69	200	163	118	9	109
2.	19	91	52	107	1	106
3.	50	109	111	11	8	3
1. zinnen met een persoonsvorm als eerste woord						
2. vraagzinnen met een persoonsvorm als eerste woord						
3. overige zinnen met een persoonsvorm als eerste woord						

Vervolgens zijn de zinnen opgevraagd met een vraagwoord als eerste, tweede of derde woord. Voor een vraagwoord worden gebruikt: de codes 40., 41. (interrogativa) en 520, 550 (vragende, eventueel voornaamwoordelijke, bijwoorden). Bovendien moeten ook opgevraagd worden de zinnen die beginnen met het woord *hoeveel* dat volgens het Codeervoorschrift als indefinitum wordt gecodeerd. Bij de aldus verkregen zinnen doet zich het probleem voor dat ook vraagwoorden van de afhankelijke

vraag eronder vallen, bijvoorbeeld *Tot wie 400 u zich moet wenden, kan ik niet zeggen* en *Hoeveel 440 ik verdien, weet ik niet*. Maar het oordeel over het geringe aantal vraagzinnen heeft naar alle waarschijnlijkheid geen betrekking op afhankelijke vragen. Daarom moest nog worden nagegaan hoeveel zinnen een afhankelijke vraag bevatten. Een tweede probleem is dat onder 40. en 41. ook de uitroepen vallen, *Wat 400 jammer!* en bijvoorbeeld *wie in Wie 400 er ook belt, doe niet open!* Ook deze zinnen moeten geschrapt worden.

Voor het bepalen van de frequenties was het dus noodzakelijk via *Oog op Tekst* alle zinnen op te vragen met de hierboven genoemde codes en het woord *hoeveel*, opdat alsnog bepaald zou kunnen worden hoe vaak er sprake was van een vraagwoord bij een niet-afhankelijke vraag.

Van de zinnen met een vraagwoord als eerste, tweede of derde woord bleek het merendeel echte vraagzinnen te zijn. Slechts een klein gedeelte betrof zinnen die beginnen met een afhankelijke vraag. Bij de overige gevallen ging het om uitroepen of zinnen van het type *Wat 400 je ook doet, het helpt niet* en *Hoe 520 eerder hoe beter*. Overigens waren er in totaal slechts vijf vraagzinnen met een vraagwoord als derde woord, bijvoorbeeld *En in welke 413 mate bent u het daar mee eens?* In schema:

	D	O	W	TO	(R)	(StG)
1.	219	411	339	135	48	87
2.	74	164	95	49	6	43
3.	45	118	65	40	-	40
4.	12	17	15	7	5	2
5.	17	29	15	2	1	1

1. vraagwoorden (codes 400, 402, 403, 410, 413, 520, 550 en *hoeveel*)
2. zinnen met een vraagwoord als eerste, tweede of derde woord
3. vraagzinnen met een vraagwoord als eerste, tweede of derde woord
4. zinnen die beginnen met een afhankelijke vraag
5. uitroepen en overige gevallen

Het totaal aantal vraagzinnen dat begint met een persoonsvorm, of een vraagwoord als eerste, tweede of derde woord kwam zo op:

D	O	W	TO	(R)	(StG)
64	209	117	147	1	146

Deze getallen werden gebruikt voor het toetsen van hypothese 15 over vraagzinnen.

Tot slot van deze paragraaf over het menselijk element nog drie hypothesen. Oordelen over de onpersoonlijke overheidstaal kunnen ook in verband gebracht worden met de C₃C-categorieën: eigennamen, interjecties of interjectivisch gebruikte woorden, en imperatieven. Vandaar:

16. eigennamen

In TO komen de codes 010, 011 en 012 minder voor dan in D, O of W.

17. interjecties

In TO komen de codes 080, 180, 287, 580, 800 en 810 minder voor dan in D, O of W.

18. imperatieven

In TO komen de codes 247, 248, 257, 258, 267, 277 en 278 minder voor dan in D, O of W.

De gegevens voor deze drie hypothesen staan vermeld in de volgende tabel:

	D	O	W	TO	(R)	(StG)
eigennamen	6794	5000	2525	953	669	284
interjecties	71	178	56	19	10	9
imperatieven	46	105	62	15	12	3

De codes 081 (interjectie in meervoudsvorm) en 268 (een imperatief van een reflexief gebruikt werkwoord met pronomen personale) kwamen niet voor. Bij het toetsen zijn de imperatieven vergeleken met het totaal en met het aantal woorden van de woordsoort waartoe ze behoren, de persoonsvorm.

6.2.2 Ingewikkeld

6.2.2.1 Bijzinnen en tussenzinnen

Hoewel het C₃C-systeem niet bedoeld is om verschijnselen betreffende de zinsstructuur in kaart te brengen, kunnen binnen het C₃C-kader toch een aantal hypothesen worden geformuleerd over de structuur van zinnen.

Naar aanleiding van het oordeel dat in TO de zinnen meer bij- en tussenzinnen bevatten kan de hypothese worden opgesteld dat er in TO meer persoonsvormen voorkomen:

19. persoonsvormen

In TO komen de codes 24., 25., 26. en 27. meer voor dan in D, O of W.

D	O	W	TO	(R)	(StG)
11432	12536	11662	3695	2609	1086

Ook zinslengte kan met zinsstructuur in verband gebracht worden. Hoe meer bij- en tussenzinnen, des te groter de kans dat de zinnen langer zijn. De zinslengte is berekend op basis van het aantal codes per zin; dit houdt dus in dat een eigennaam bestaande uit meer dan één woord, als één woord is geteld.

20. zinslengte

In TO zijn de zinnen over het algemeen langer dan in D, O of W.

De gegevens over zinslengte staan hieronder vermeld. (Het begrip standaarddeviatie wordt toegelicht op pagina 196.)

	D	O	W	TO	(R)	(StG)
aantal zinnen	6631	6812	5752	1955	1416	539
gemiddelde zinslengte	18,30	18,41	21,08	24,64	25,05	23,58
standaarddeviatie	9,59	12,08	13,22	12,68	12,77	12,38

Naast deze hypothese zijn er nog andere op te stellen op basis van het voorkomen van woordsoorten die een bijzin inleiden. Bijzinnen kunnen worden ingeleid door *relativa*, *betrekkelijke* (*voornaamwoordelijke*) *bijwoorden* en *onderschikkende voegwoorden*. Vandaar de volgende hypothesen:

21. *relativa*

In TO komen de codes 420, 422, 423, 430 en 433 meer voor dan in D, O of W.

22. *betrekkelijke (voornaamwoordelijke) bijwoorden*

In TO komen de codes 530 en 560 meer voor dan in D, O of W.

23. *onderschikkende voegwoorden*

In TO komen de codes 710 en 730 meer voor dan in D, O of W.

Naar aanleiding van het oordeel betreffende beknopte bijzinnen kan een hypothese opgesteld worden over het voorkomen van code 650. Deze code wordt gebruikt voor *te* bij een infinitief en een voorzetsel met een *te*+infinitiefconstructie als rest, bijvoorbeeld *teneinde 650 hieraan gestalte te 650 kunnen geven*. Na een ander voorzetsel dan *te* met code 650 komt er dus altijd *te 650*. De hypothese is beperkt tot *te 650*, om te vermijden dat eventuele verschillen in frequenties van andere voorzetsels met code 650 een storende invloed uitoefenen.

24. *'te' 650*

In TO komt het woord *'te'* met code 650 meer voor dan in D, O of W.

Tenslotte kan het oordeel over ingewikkelde zinsbouw ook in verband gebracht worden met *nevenschikkende voegwoorden* of *voegwoorden van vergelijking*. Vandaar:

25. *nevenschikkende voegwoorden*

In TO komen de codes 700 en 740 meer voor dan in D, O of W.

26. *voegwoorden van vergelijking*

In TO komt code 720 meer voor dan in D, O of W.

De gegevens voor het toetsen van de hypothesen 21 tot en met 26 zijn ondergebracht in de volgende tabel:

	D	O	W	TO	(R)	(StG)
21. relativa	1288	1465	1438	402	308	94
22. betrekkelijke (voornaamwoordelijke) bijwoorden	451	463	466	172	125	47
23. onderschikkende voegwoorden	1715	2266	2080	886	594	292
24. 'te' 650	1245	1386	1214	550	392	158
25. nevenschikkende voegwoorden	3294	4106	4190	1075	800	275
26. voegwoorden van vergelijking	643	823	903	253	183	70

6.2.2.2 Tangconstructies

Het syntactische verschijnsel 'tangconstructie', het niet bij elkaar staan van zinsdelen die bij elkaar horen, kan via het C_3C -systeem niet of nauwelijks direct onderzocht worden.

In het C_3C -systeem zijn slechts drie codes die met dit verschijnsel te maken hebben: 610 (tweede deel van gesplitste voornaamwoordelijke bijwoorden) *Daar 540 doe ik het mee 610*; 620 (niet-verbaal deel van samengestelde werkwoorden) *Hij belde 255 mij op 620*, en 630 (tweede deel van gescheiden voorzetselgroepen) *tegen 600 de voorschriften in 630*. Naar aanleiding van het oordeel over het veelvuldig voorkomen van tangconstructies in de taal van 'Den Haag' zou men mogen verwachten, dat de afstand tussen deze codes en de bijbehorende woorden in TO over het algemeen groter is dan in D, O of W.

Voor het toetsen van hypothesen hierover zijn via *Oog op Tekst* de zinnen opgevraagd met de codes 610, 620 en 630. De frequenties voor 610 en 620 liggen aanzienlijk hoger dan zoals vermeld in *Woordfrequenties*. Dit komt doordat in *Woordfrequenties* alleen is weergegeven hoe vaak deze codes voorkomen zonder bijbehorend voornaamwoordelijk bijwoord of werkwoord. Ook bij 630 bleek een klein verschil, maar dit is te wijten aan enkele foute coderingen.

Vervolgens zijn met de hand de woorden geteld tussen de woorden met deze codes en de bijbehorende woorden. Op de volgende pagina staan de frequenties van de codes en de gemiddelde afstanden met standaarddeviatie.

		D	O	W	TO	(R)	(StG)
610	1.	339	452	260	35	25	10
	2.	2,16	2,26	2,10	1,43	1,16	2,10
	3.	2,27	2,02	1,84	1,17	1,14	0,99
620	1.	592	619	587	206	146	60
	2.	3,00	2,58	2,76	2,50	2,45	2,60
	3.	3,30	2,85	3,11	3,34	3,37	3,29
630	1.	66	97	68	8	6	2
	2.	1,47	1,85	1,65	1,13	1	1,5
	3.	1,32	1,93	1,35	0,35	0	0,71

1. frequentie

2. gemiddelde afstand (540-610; 2...-620; 600-630)

3. standaarddeviatie

Deze getallen werden gebruikt bij het toetsen van de volgende hypothesen:

27. afstand tussen delen van voornaamwoordelijke bijwoorden

In TO is de afstand tussen code 610 en het bijbehorende woord over het algemeen groter dan in D, O of W.

28. afstand tussen delen van samengestelde werkwoorden

In TO is de afstand tussen code 620 en het bijbehorende woord over het algemeen groter dan in D, O of W.

29. afstand binnen voorzetselgroepen

In TO is de afstand tussen code 630 en het bijbehorende woord over het algemeen groter dan in D, O of W.

Van de verschijnselen die in hoofdstuk 3 genoemd worden onder 'tangconstructies', kunnen slechts enkele geherformuleerd worden tot hypothesen, namelijk het verschijnsel dat voegwoorden naast elkaar staan en het voorkomen van een bepaald type klemconstructie met een daaraan verwant verschijnsel.

Het verschijnsel 'samentreffen van voegwoorden' dat, omdat een bijzin in een hoofd- of bijzin wordt opgenomen, een tangconstructie oplevert (zoals in deze zin) is als volgt onderzocht. Via *Oog op Tekst* zijn de

zinnen opgevraagd met de codes 700 of 710, direct gevolgd door code 7.. Bij de combinaties met 700 bleek alleen van een tangconstructie sprake wanneer code 710, en in een enkel geval code 730, erop volgde. Buiten beschouwing bleven de combinaties: 700-700 *Hij deed het op verzoek van ouders en 700 of 700 leerkrachten*; 700-720 *Hij is benoemd als violist en 700 als 720 dirigent* en 700-740 *Want 700 noch 740 over Jansen noch over Pietersen is iets bekend*. Verder zijn niet meegeteld die combinaties 700-710 waarin het woord met code 700 alleen als verbindingswoord dienst doet tussen twee bijzinnen: *Hij zei dat hij ziek was en 700 dat 710 hij in bed bleef*. Hier is geen sprake van inbedding in een hoofd- of bijzin. Dit is bijvoorbeeld wel het geval in *Je moet minder eten, want 700 als 710 je zo doorgaat, word je te dik*. Alleen zinnen van het laatste type zijn meegeteld als het ging om combinaties 700-7.. Bij de mogelijkheden 710-7.. is alleen de combinatie 710-710 meegeteld. Bijvoorbeeld: *Het lijkt mij beter dat 710 wanneer 710 hij aanbelt, er niet open gedaan wordt*. Hieronder de aantallen in schema.

	D	O	W	TO	(R)	(StG)
1.	86	155	114	31	20	11
2.	14	24	20	5	3	2
3.	26	66	59	22	15	7
4.	46	65	35	4	2	2
5.	12	14	24	25	14	11
6.	10	7	15	17	8	9
7.	2	7	9	8	6	2

1. totaal aantal 700-7..
2. combinaties 700-700, 700-720, 700-740
3. gevallen waarin 700 alleen hoofd- of bijzinnen koppelt
4. combinatie van 700-710 (of 730) met tangconstructie
5. totaal aantal 710-7..
6. combinatie van 710-710 met tangconstructie
7. andere combinaties met 710

De getallen achter 4 en 6 werden gebruikt voor het toetsen van de volgende hypothese:

30. voegwoord plus voegwoord

In TO komt de combinatie 7..-7.. meer voor dan in D, O of W.

Bij het toetsen van het oordeel over het veelvuldig voorkomen van klemconstructies ("de in deze nota over deze kwestie gegeven antwoorden"), en vooral van lange klemconstructies, is gebruik gemaakt van een belangrijk kenmerk van deze constructie, namelijk het feit dat er dikwijls vlak voor het substantief een deelwoord of een infinitief staat. De hypothese is beperkt tot dit type klemconstructie.

Via *Oog op Tekst* zijn de zinnen opgevraagd waarin een woord met code 20., 21., 22. of 23. (een deelwoord of infinitief) direct gevolgd wordt door een woord met code 0.. (een substantief). Daarna is met de hand geteld het aantal woorden tussen het woord dat behoort bij het substantief, en het deelwoord of de infinitief.

Het woord dat behoort bij het substantief moet code 3.. of 4.. hebben; meestal is het een lidwoord. Een voorbeeld: *het 370 reeds eerder gedane 217 voorstel 000* is geteld als een klemconstructie van twee woorden. Wanneer in plaats van code 3.. of 4.. de woordgroep 'een aantal', 'een paar', 'een soort' of een equivalent daarvan voorkwam, is er geteld vanaf deze woordgroep. Dus *een 450 aantal 000 te ordenen 210 pagina's 001* is een klemconstructie van één woord. In *twee 460 van de drie 470 onderzochte 217 variabelen 001* zou men eventueel kunnen tellen vanaf *twee 460*, maar in twijfelgevallen is steeds gekozen voor het dichtstbijzijnde woord met code 3.. of 4..

Op de plaats van 0.. kan ook een woord staan met code 010 of 020: *de 370 in aanleg zijnde 233 Molenstraat 010; een 450 al jaren niet meer opgelegde 217 maximum 020 straf 000* (klemconstructies van respectievelijk twee en vier woorden). Verder zij opgemerkt dat een klemconstructie in een klemconstructie niet apart is geteld, dus zeven woorden in *de 370 door de 370 tot de k.v.p.-fractie behorende 203 leden 001 gedane 217 suggestie 000*.

Van de zinnen die via *Oog op Tekst* werden verkregen, moest een groot aantal buiten beschouwing blijven. In een aantal gevallen was er bij de opeenvolging van deelwoord of infinitief en substantief geen sprake van een klemconstructie. Bijvoorbeeld *Door de afwezigheid van het vaderimago ontstaat, zoals is aangetoond 206, haat 000* of *Hij heeft voorgesteld 216 wachtlijsten 001 in te voeren*. Deze gevallen staan in de tabel op de volgende pagina vermeld achter I. Achter II staan de constructies waarin geen bijbehorend woord met 3.. of 4.. aanwezig was, bijvoorbeeld *nieuwe 103 nog niet eerder gehoorde 217 gevallen 001*.

Onder I en II staan de aantallen klemconstructies en de gemiddelde afstand met standaarddeviatie.

	D	O	W	TO	(R)	(StG)
I	79	68	67	22	16	6
II	201	194	224	125	99	26
1.	717	617	729	540	421	119
2.	1,11	1,27	1,04	1,55	1,61	1,34
3.	1,69	1,98	1,70	2,20	2,28	1,87

I deelwoord of infinitief plus substantief, zonder klemconstructie

II deelwoord of infinitief plus substantief, zonder bijbehorend woord met 3.. of 4..

1. aantal klemconstructies
2. gemiddelde afstand
3. standaarddeviatie

Deze getallen werden gebruikt voor het toetsen van de volgende hypothesen:

31. klemconstructies met deelwoord of infinitief

In TO is het aantal klemconstructies met deelwoord of infinitief groter dan in D, O of W.

32. aantal woorden per klemconstructie

In TO is het aantal woorden per klemconstructie met deelwoord of infinitief over het algemeen groter dan in D, O of W.

Tot slot van deze paragraaf over tangconstructies nog één hypothese.

Nauw verwant aan de klemconstructie is het in hoofdstuk 3 genoemde verschijnsel van de opeenhoping van lidwoorden en voorzetsels, bijvoorbeeld "Er kleven bezwaren aan de door de minister voorgestelde maatregelen." Voor het toetsen van een hypothese over dit verschijnsel zijn via *Oog op Tekst* opgevraagd de zinnen waarin voorkomt de woordsoortvolgorde: voorzetsel-lidwoord-voorzetsel-lidwoord, of de omgekeerde volgorde. Voor de voorzetsels is alleen opgevraagd code 60. Dit houdt in dat uitgeschakeld worden zinnen als ... *om 650 de 370 door 600 de 370 betrokkenen zelf ervaren malaise te doorbreken*. Onder 60. valt wel bijvoorbeeld *tijdens 600 de over de zaak gevoerde discussie*. De bepaalde

lidwoorden vallen in het codeersysteem onder code 37., bijvoeglijke demonstrativa. Bij het opvragen van zinnen kwamen ook andere woorden met deze code in aanmerking. Onder de hypothese staan de getallen die gebruikt zijn bij de toetsing.

33. ketens van voorzetsels en lidwoorden

In TO komen de ketens 'voorzetsel-lidwoord-voorzetsel-lidwoord' en 'lidwoord-voorzetsel-lidwoord-voorzetsel' meer voor dan in D, O of W.

D	O	W	TO	(R)	(StG)
25	21	26	52	44	8

Hierbij zij nog opgemerkt, dat er slechts in 16 gevallen sprake was van zinswendingen die niet vallen onder de klemconstructie zoals beschreven op pagina 123, bijvoorbeeld *Door 600 het 370 uit 600 het 370 veld zenden van twee spelers, wist hij zijn positie te verstevigen*. Overigens kwam de omgekeerde volgorde (lidwoord-voorzetsel-lidwoord-voorzetsel) maar één keer voor, en wel in subcorpus R.

6.2.2.3 Andere zinsbouwverschijnselen

De elliptische constructies waarover onder dit hoofdje in hoofdstuk 3 werd gesproken kunnen niet via het C₃C-systeem onderzocht worden. De opmerkingen over onhandige of dubbele ontkenningen kunnen eveneens moeilijk tot een hypothese worden geherformuleerd. Wel kan naar aanleiding van het oordeel dat de lezer van overheidsteksten te weten komt 'hoe iets niet is', de volgende hypothese worden opgesteld.

34. ontkennende woorden

In TO komen ontkennende woorden meer voor dan in D, O of W.

Tot de ontkennende woorden zijn gerekend: *allerminst, evenmin, geen* (en de vormen *genen, gener, generlei*), *geensdeels, geenszins, niemand, niet, niets, nimmer, niks, noch* en *nooit*. In de tabel op de volgende pagina staan de frequenties van de woorden die in de steekproeven voorkomen.

		D	O	W	TO	(R)	(StG)
allerminst	447/457	5	7	9	1	-	1
evenmin	500	3	6	11	3	-	3
geen	440/450	276	299	277	79	49	30
geenszins	500	1	1	6	4	3	1
niemand	440	13	34	17	1	1	-
niet	500	800	1194	941	349	252	97
niets	440/450	74	88	51	4	2	2
niks	440	4	8	3	-	-	-
nimmer	510	3	3	6	2	-	2
noch	700/740	10	16	19	8	6	2
nooit	510	40	64	35	1	1	-
		1229	1720	1375	452	314	138

Maar het onderzoeken van alleen deze woorden is nog niet voldoende. Een zin kan ook een ontkennend karakter dragen wanneer er woorden in voorkomen die beginnen met 'niet-', 'non-' of 'on-'. Helaas is bij woorden met 'on-', op basis van de frequentielijsten (dus zonder context) lang niet altijd duidelijk of er werkelijk sprake is van een ontkenning. Bij een woord als 'ontologisch' is dit uiteraard geen probleem; het gaat om het ontkennende prefix. Bij een aantal andere woorden is ook zonder meer duidelijk dat het niet om ontkenningen gaat, bijvoorbeeld 'onmiddellijk', 'onlangs', 'onmiskenbaar'. Maar bij andere woorden zou de context moeten worden betrokken om hierover een beslissing te nemen, bijvoorbeeld 'onafhankelijkheid', 'onbewust', 'oneindig'. En dan nog is een beslissing tot op zekere hoogte arbitrair.

Een poging om al zonder context de 'on-woorden' te selecteren die niet of niet per se ontkenningen inhouden, leverde de aantallen op die in het volgende schema vermeld staan achter 3. Deze aantallen zijn bij het toetsen van de hypothese buiten beschouwing gelaten. Bij de woorden met 'on-' doen alleen de woorden mee die al zonder context een duidelijk ontkennend karakter hebben. Op de volgende pagina staan de aantallen voor de woorden met 'niet-', 'non-' en 'on-'.

	D	O	W	TO	(R)	(StG)
1.	11	24	24	14	12	2
2.	293	331	336	84	60	24
3.	184	154	136	38	28	10
1. woorden met 'niet-' of 'non-'						
2. woorden met 'on-'						
3. woorden met 'on-' die niet negatief zijn						

De getallen achter 1 en 2 zijn gebruikt voor het toetsen van de volgende hypothese:

35. woorden met 'niet-', 'non-' of 'on-'

In TO komen woorden met 'niet-', 'non-' of 'on-' meer voor dan in D, O of W.

6.2.2.4 Moeilijke woorden

Oordelen hierover zijn moeilijk te onderzoeken bij gebrek aan een werk-bare omschrijving van de term 'moeilijke woorden'. Hetzelfde geldt voor vreemde woorden en afkortingen, die in het codeersysteem C₃C niet van een aparte code worden voorzien.

Het enige wat naar aanleiding van deze oordelen onderzocht kan worden is de gemiddelde woordlengte. Hierbij is uitgegaan van de vooronderstel-ling dat moeilijke woorden over het algemeen genomen langer zijn dan gemakkelijke woorden. (Zie o.a. *Van Haanwermeiren* 1975:43.) Helaas kon de gemiddelde woordlengte alleen in letters worden berekend, en niet in lettergrepen zoals in leesbaarheidsonderzoek gebruikelijk is. In de volgende tabel staan de uitslagen vermeld.

	D	O	W	TO	(R)	(StG)
aantal woorden	120313	123873	119979	48257	35300	12957
gemiddelde woordlengte	5,39	5,28	5,32	5,57	5,56	5,59
standaard-deviatie	3,57	3,55	3,47	3,72	3,65	3,89

De hypothese luidt als volgt:

36. woordlengte

In T0 zijn de woorden over het algemeen langer dan in D, O of W.

6.2.3 Wijdlopig

De oordelen die onder dit hoofdje zijn genoemd, lenen zich bijzonder moeilijk voor herformulering binnen het kader van C₃C of *Oog op Tekst*.

De enige opmerking die nader bekeken kan worden, is dat een tekst meer bijvoeglijke naamwoorden en bijwoorden bevat, naarmate er meer 'overbodige bepalingen' of 'stijlparasieten' (zie pagina 41) in voorkomen. Bij het opstellen van hypotheses hierover zijn alleen de 'gewone' adjectieven, de adverbiaal gebruikte adjectieven en de 'gewone' bijwoorden betrokken. De andere coderingen onder adjectieven en bijwoorden zijn bedoeld voor woordsoorten die niet met wijdlopigheid in verband gebracht kunnen worden. De hypotheses luiden als volgt:

37. adjectieven

In T0 komen de codes 100, 102, 103, 104, 105, 106, 107 en 109 meer voor dan in D, O of W.

D	O	W	T0	(R)	(StG)
6468	7373	7320	2566	1886	680

38. bijwoorden

In T0 komen de codes 150, 153, 154, 157, 159, 500, 503, 504, 507 en 509 meer voor dan in D, O of W.

D	O	W	T0	(R)	(StG)
7564	8619	8026	2201	1555	646

6.2.4 Traditioneel

De oordelen over archaïsch taalgebruik kunnen alleen onderzocht worden door na te gaan welke codes hierop betrekking kunnen hebben. De opmerkingen over clichématig taalgebruik lenen zich niet voor kwantitatief-stilistisch onderzoek, bij gebrek aan een nauwkeurige definitie.

Als archaïsch kunnen beschouwd worden de woorden met de codes voor 'overige vormen van een persoonsvorm', voornamelijk de aanvoegende wijs. Vandaar de volgende hypothese:

39. aanvoegende wijs

In TO komen de codes 249, 259 en 279 meer voor dan in D, O of W.

D	W	O	TO	(R)	(StG)
29	31	60	35	27	8

Code 269, een reflexief gebruikt werkwoord in de aanvoegende wijs, kwam niet voor. De getallen zijn niet alleen vergeleken met het totaal aantal woorden, maar ook met het aantal woorden dat als persoonsvorm is gecodeerd.

Ook bij de volgende drie hypothesen is er, naast een toetsing op het totaal, een vergelijking gemaakt met het aantal woorden van de woordsoorten waartoe de desbetreffende woorden behoren (zie hierover verder pagina 135).

Behalve de aanvoegende wijs kunnen ook naamvalsvormen van substantieven worden betrokken bij hypothesen over archaïsch taalgebruik (derde cijfer 2 of 3). Maar niet alle woorden met code 0.2 zijn archaïsch. Enkele voorbeelden: 's 372 middags 002; Jan's 012 jas. Code 003, substantieven in overige verbogen vormen, geeft op dit punt geen problemen. Bij het opstellen van een hypothese over substantieven in een naamvalsvorm is deze categorie in haar geheel betrokken, maar van categorie 002 zijn de woorden die een deel van de dag aangeven, 's morgens enz., of een seizoen, 's zomers enz., buiten beschouwing gelaten. (Woorden met code 012 zijn niet als archaïsch te beschouwen.) Deze woorden konden gemakkelijk worden opgespoord op basis van de frequentielijsten, alfabetisch gerangschikt naar code. Op de volgende pagina staan de resultaten:

	D	O	W	TO	(R)	(StG)
1.	44	31	35	4	3	1
2.	10	10	22	4	3	1
3.	84	98	134	107	80	27
1. woorden met code 002						
2. woorden met code 002, zonder aanduidingen voor delen van de dag of seizoenen						
3. code 003						

De getallen in rij 2 en 3 werden gebruikt voor het toetsen van de volgende hypothese:

40. substantieven in verouderde naamvalsvormen

In TO komen code 003 en een deel van code 002 meer voor dan in D, O of W.

Verder zijn voor het derde cijfer 2 of 3 de pronomina het onderzoeken waard. De categorie adjectieven met ..2 of ..3 bevat nauwelijks archaïsche vormen; hetzelfde geldt voor bijwoorden met code 5.3. Zie hierover pagina 53 en 62. Maar ook bij de pronomina kan niet zonder meer gesteld worden dat ..2 of ..3 betrekking heeft op een archaïsche vorm. Enkele voorbeelden: *onze* 333; *des* 362 *te beter*; *degene* 363; *iets dergelijks* 372; *welke?* 413. Daarom is, op basis van de frequentielijsten alfabetisch gerangschikt naar code, bij de personalia en de possessiva gekeken of, en zo ja hoeveel, verouderde naamvalsvormen voorkwamen. Hierbij zij nog vermeld dat *des* met code 362 (*des te beter*) buiten beschouwing is gelaten, evenals 's 372. Verder valt *den* in namen onder categorie 010 en krijgen vormen als *bij deze* of *dezelve* als derde cijfer een 0; deze woorden vallen dus niet onder verouderde naamvalsvormen zoals hier omschreven. In het schema op de volgende pagina is per code aangegeven welke woorden zijn meegeteld. Hierbij kan nog worden opgemerkt dat de frequenties voor *der* 372 meer dan de helft van het totaal uitmaken: D 60; O 53; W 225; TO 21; R 19; StG 2.

	D	O	W	TO	(R)	(StG)
302 (haars, hunner, onzer)	-	4	1	-	-	-
303 (onzent)	-	-	1	-	-	-
323 (zijne, hunne)	-	1	2	-	-	-
332 (mijns, zijner, zijns, hunner, huns, onzer)	3	2	8	9	8	1
333 (mijne, mijner, uwe, zijne, hunne, hunner, hunnen, sijnen, zijnen)	2	4	5	2	2	-
342 (zichzelfs)	-	1	-	-	-	-
343 (zichzelve)	-	1	-	-	-	-
353 (zelve)	1	-	-	2	2	-
362 (diens, dies, dezes)	12	8	7	5	4	1
363 (dezen, dien)	2	1	2	5	3	2
372 (der, des, dezer, diens, dier)	74	64	243	23	21	2
373 (den, der, dien, dezer)	9	9	18	5	5	-
402 (wiens)	-	3	1	-	-	-
422 (wiens, welks, wier)	13	11	14	2	1	1
452 (ener, eether)	-	-	2	-	-	-
	116	109	304	53	46	7

Deze getallen werden gebruikt voor het toetsen van de volgende hypothese:

41. pronomina in verouderde naamvalsvormen

In TO komen pronomina in verouderde naamvalsvormen meer voor dan in D, O of W.

Ook de woorden met code 603, bijvoorbeeld *ten 603 aanzien van*, zijn tot de archaïsmen gerekend. Vandaar:

42. verbogen voorzetsels

In TO komt code 603 meer voor dan in D, O of W.

D	O	W	TO	(R)	(StG)
132	134	194	220	184	36

Tot slot kan er, wat archaisch taalgebruik betreft, gekeken worden naar de woorden die onder 'boekentaal' staan opgesomd op pagina 42, voor zover ze nog niet verwerkt zijn in vorige hypothesen. De woorden die daar staan vermeld, zijn eigenlijk niet binnen een systematisch verband onder te brengen. Het is een samenraapsel van woorden vermeld in literatuur over de taal van 'Den Haag'. Hieronder volgen per woord de frequenties.

	D	O	W	TO	(R)	(StG)
alsmede	8	3	7	15	9	6
alsook	-	-	3	-	-	-
betreffende 103	-	1	-	-	-	-
hetgeen	14	16	32	17	13	4
hetwelk	-	1	1	-	-	-
middels	-	-	1	1	1	-
onderhavig(e)	1	1	1	21	16	5
uiteraard	7	17	17	16	15	1
waar 710	-	1	-	1	1	-
wat ... betreft	10	22	27	26	20	6
welk(e) 42./43.	16	13	44	42	35	7
	56	75	133	139	110	29

Voor *wat ... betreft* zijn via *Oog op Tekst* de zinnen opgevraagd. Het woord *mitsgaders* kwam niet voor. Bij *betreffende* gaat het om het bijvoeglijk gebruik: *het betreffende 103 stuk*. Code 710 achter *waar* wil zeggen dat het alleen gaat om het voegwoordelijk bijwoord, zoals bedoeld op pagina 42. En bij *welk(e)* is alleen het betrekkelijk voornaamwoord bedoeld.

Deze getallen werden gebruikt voor het toetsen van de laatste hypothese:

43. enkele archaïsmen

In TO komen de woorden *alsmede*, *alsook*, *betreffende 103*, *hetgeen*, *hetwelk*, *middels*, *onderhavig(e)*, *uiteraard*, *waar 710*, *wat ... betreft* en *welk(e) 42./43.* meer voor dan in D, O of W.

6.3 Toetsing

In deze paragraaf staat beschreven hoe de hypothesen over de taal van 'Den Haag' zijn getoetst (6.3.1) en wat de uitkomsten zijn (6.3.2).

6.3.1 Wijze van toetsing

Voor een toelichting op de manier van toetsen neem ik als voorbeeld de eerste hypothese over de naamwoordstijl. Deze hypothese hield in dat er in teksten van de overheid (TO) meer substantieven staan dan in dagbladen (D), opiniebladen (O) of populair-wetenschappelijke boeken (W). Deze hypothese bestaat eigenlijk uit vier sub-hypothesen. TO wordt immers vergeleken met D, met O en met W. Bovendien wordt onderzocht of er binnen TO verschillen zijn aan te wijzen tussen R en StG. Er is dus sprake van twee typen hypothesen:

A. Er is verschil tussen TO en D, O of W wat een bepaald taalverschijnsel betreft. Hierbij wordt een uitspraak gedaan over de verwachte richting van het verschil, met andere woorden: of het verschijnsel in TO meer of minder voorkomt dan in D, O of W.

B. Er is verschil tussen R en StG, maar er wordt geen uitspraak gedaan over de verwachte richting van het verschil.

De nul-hypothese (H_0) luidt dan dat er geen verschil is. In dit geval kiezen we een tweezijdig alternatief (H_1), omdat bij hypothesen van het type A zowel significante afwijkingen in verwachte richting als in niet-verwachte richting interessant zijn, en omdat bij hypothesen van het type B geen uitspraak wordt gedaan over de richting van het verschil.

Wanneer we toetsen, houdt dit in dat we mogen aannemen dat het alternatief is bevestigd, als we kunnen aantonen dat de nul-hypothese verworpen mag worden. Bij het toetsen van hypothesen bepaalt men eerst hoe betrouwbaar de toetsing moet zijn, of anders gezegd: men bepaalt vooraf hoe groot de kans mag zijn, dat men een nul-hypothese ten onrechte verworpt. Wanneer een onderzoeker bijvoorbeeld zegt: Ik wil werken met de kans dat ik 1 op de 20 keer ten onrechte een nul-hypothese verwerp (en zodoende dus ten onrechte aanneem dat het alternatief is bevestigd), dan heeft hij een onbetrouwbaarheidsdrempel α gekozen van vijf procent, ofwel $\alpha = 0,05$. Maar een onderzoeker kan zijn kansen op onjuiste conclusies kleiner maken door bijvoorbeeld als drempel te nemen 1 op de 100 keer of zelfs 1 op de 1000 keer ($\alpha = 0,01$ of $0,001$). Het zal duidelijk zijn dat de uitspraken naar aanleiding van de uitkomsten voorzichtiger worden naarmate de onbetrouwbaarheidsdrempel

lager is. Bij het toetsen van hypothesen over de taal van 'Den Haag' is als waarde voor α genomen 0,001. Met andere woorden, de kans dat ten onrechte een nul-hypothese wordt verworpen, en het alternatief dus ten onrechte wordt bevestigd, is 1 op 1000.

Wanneer nu de hypothese wordt getoetst dat er in TO meer substantieven voorkomen dan in bijvoorbeeld D, moet eerst aangetoond worden dat de nul-hypothese (namelijk dat TO evenveel substantieven bevat als D) verworpen mag worden. En bovendien moet er afwijking in verwachte richting zijn.

De gegevens die voor het toetsen van deze hypothese nodig zijn, kunnen ondergebracht worden in een zogenaamde '2 x 2-tabel' (in het schema hieronder de eerste twee kolommen). In deze tabel is in de eerste rij (subgroep) weergegeven hoeveel substantieven, dus de woorden met codes zoals aangegeven in hypothese 1, zijn aangetroffen in corpus D uit *Woordfrequenties* en in corpus TO. In de tweede rij (restant) staat het aantal overige woorden uit de beide corpora. Onder deze vier getallen staan ook de steekproef-fracties vermeld. (Fracties liggen tussen 0 en 1; de fractie vermenigvuldigd met 100, geeft het percentage weer.) Onder deze tabel ziet u de getallen van de kolommen, van subgroep en restant, bij elkaar opgeteld. Daar staat dus de omvang van de corpora D en TO. In de derde kolom vindt u de getallen van de twee subgroepen en de twee restanten opgeteld. En rechts onder de tabel staan de beide corpora bij elkaar opgeteld. Dit laatste getal is dus de uitkomst van de optellingen zowel verticaal als horizontaal van de getallen buiten de 2 x 2-tabel

	D	TO	
Subgroep	24299 0,2020	11206 0,2322	35505 Afwijking in verwachte richting
Restant	96014 0,7980	37051 0,7678	133065 Overschrijdingskans: 0,00000000
	120313	48257	168570

Naast deze tabel staan twee gegevens die verkregen zijn na toepassing van een chi-kwadraattoets. De beoordeling van de te toetsen hypothese geschiedt met behulp van een toetsingsgrootte die uit de waarnemingen wordt berekend. Door middel van een chi-kwadraattoets valt na te gaan of er een significant verschil bestaat tussen de populatie-fracties in de subgroep. Voor een duidelijke uiteenzetting van deze toets en een

rechtvaardiging van het gebruik ervan, wordt verwezen naar *Hemelrijk* (1977).

Wanneer een nul-hypothese juist is, is er geen significant verschil tussen beide populatie-fracties uit de subgroep. Via statistische bewerkingen valt uit te rekenen hoe groot de kans is, onder aanneming van de nul-hypothese, dat men boven de waargenomen waarde van de toetsingsgrootte uitkomt. Men noemt dit de overschrijdingskans. Deze kans wordt dan vergeleken met de gekozen onbetrouwbaarheidsdrempel (in ons geval 0,001). Is deze kans kleiner dan de onbetrouwbaarheidsdrempel, dan verwerpen we H_0 ; is deze kans groter, dan verwerpen we H_0 niet. De overschrijdingskans zegt iets over de sterkte van de bewering. Wanneer de gevonden overschrijdingskans veel kleiner is dan 0,001, dan zal men met des te meer overtuiging de nul-hypothese mogen verwerpen.

De gegevens naast de voorbeeld-tabel mogen nu als volgt geïnterpreteerd worden: Er is een significante afwijking in verwachte richting. Dus in TO komen meer substantieven voor dan in D. Uiteraard kan er naast zo'n tabel ook staan 'afwijking in niet-verwachte richting'. Wanneer de overschrijdingskans dan kleiner zou zijn dan 0,001, zou geconcludeerd moeten worden dat TO minder substantieven bevat dan D. Maar, wanneer de overschrijdingskans bijvoorbeeld 0,021 is, dan mag de nul-hypothese niet verworpen worden. Want deze kans is groter dan de onbetrouwbaarheidsdrempel van 0,001. Met andere woorden, er mag worden aangenomen dat de hypothese is bevestigd wanneer er staat 'afwijking in verwachte richting' en wanneer de overschrijdingskans kleiner is dan 0,001.

Voor de hypothesen waarin R en StG vergeleken worden, en waar als nul-hypothese geldt dat er geen verschil bestaat tussen R en StG, mag aangenomen worden dat de hypothese dat er wel verschil bestaat, wordt bevestigd wanneer de overschrijdingskans kleiner is dan 0,001.

Bij het toetsen zijn, met uitzondering van de hypothese over de lijdende vorm (8) en de hypothese over vraagzinnen (15), steeds de te onderzoeken woorden afgezet tegen het totaal aantal woorden. Bij een aantal hypothesen zou dit een vertekening kunnen geven. Bijvoorbeeld in TO zouden er meer woorden op '-ing' (hypothese 7) kunnen voorkomen omdat er meer substantieven zijn, terwijl het aantal '-ing'-woorden in TO relatief gezien niet groter hoeft te zijn. Daarom zijn bij enkele hypothesen de te onderzoeken woorden ook afgezet tegen de categorieën woorden waartoe ze behoren (in dit geval de woorden met codes 000 en 001). De hypothesen waarvoor dit geldt zijn: 7, 11 (verwijzingen naar een eerste persoon), 13 (men), 14 (u), 18 (imperatieven), 39 (aanvoegende wijs), 40 (substantieven in verouderde naamvalsvormen), 41 (pronomina in ver-

ouderde naamvals vormen) en 42 (verbogen voorzetsels). In enkele gevallen bleek die toetsing een ander resultaat op te leveren dan wanneer met het totaal aantal woorden werd vergeleken. Zie de 'andere uitkomsten' op pagina 139. In de Discussie wordt hierop nader ingegaan.

Wel is nog overwogen om bij verschijnselen die meer met zinsbouw te maken hebben, als maat het aantal zinnen te nemen. Het is interessant om te toetsen of er in T0, gemiddeld per zin, significant meer persoonsvormen of voegwoorden voorkomen. Maar hiervoor ontbraken de gegevens. Daarvoor is namelijk nodig een gedetailleerd overzicht van het aantal persoonsvormen of voegwoorden per zin (precies zoals bij de hypothese over bijvoorbeeld zinslengte, waar deze gegevens wel bekend waren). Het gemiddelde aantal persoonsvormen of voegwoorden per zin valt gemakkelijk te berekenen, omdat het aantal zinnen bekend is. Maar er kan dan geen uitspraak gedaan worden over al dan niet significante verschillen tussen T0 en D, O of W.

Tot slot nog een opmerking over de hypothesen waarbij de gegevens niet konden worden ondergebracht in een 2 x 2-tabel. Het betreft hier de hypothesen over zinslengte (20), afstand in tangconstructies (27, 28, 29, 32) en woordlengte (36). Als voorbeeld nemen we hier de hypothese over zinslengte.

Met behulp van betrouwbaarheidsintervallen voor de 'verwachting' van het aantal woorden per zin kan nagegaan worden of de populaties significant verschillen. ('verwachting' is een getal dat betrekking heeft op de populatie; op basis van het gemiddelde van de steekproef kan dit getal worden geschat.) Het betrouwbaarheidsinterval is voor elke populatie apart bepaald met een onbetrouwbaarheidsdrempel van 0,005. Wanneer wij twee intervallen vergelijken is de kans dus één procent dat een nul-hypothese ten onrechte wordt verworpen.

Het betrouwbaarheidsinterval voor de verwachting wordt verkregen door het gemiddelde te nemen plus of min 2,58 maal de standaarddeviatie. Voor het bepalen van dit interval (voor de verwachting van een normale verdeling met onbekende variantie) wordt weer verwezen naar *Hemelrijk* (1977). Wanneer nu de intervallen voor de gemiddelde zinslengte van T0 en bijvoorbeeld D elkaar overlappen, mag geconcludeerd worden dat T0 en D qua zinslengte niet significant verschillen. Wanneer er geen overlapping in intervallen optreedt, is er wel een significant verschil.

In Bijlage 3, pagina 177 tot 219, staan per hypothese de getallen vermeld die gebruikt zijn bij de toetsing. Nu eerst een overzicht van de resultaten.

6.3.2 Overzicht van de resultaten

In deze laatste paragraaf van dit hoofdstuk staan de uitkomsten van de toetsing. De hypothesen worden hier in trefwoorden aangeduid. De (+) of (-) na het trefwoord geeft aan of de hypothese luidt dat in TO het desbetreffende taalverschijnsel meer of minder voorkomt dan in D, O of W.

In het overzicht op de drie pagina's hierna worden de volgende afkortingen en symbolen gebruikt:

- TO teksten van de overheid
- R teksten van de regering uit TO
- StG teksten van de Staten-Generaal uit TO
- D dagbladen
- O opiniebladen
- W populair-wetenschappelijke boeken
- - geen significant verschil tussen R en StG
- + wel significant verschil tussen R en StG; het in de hypothese genoemde verschijnsel komt meer voor in de teksten waaronder de + staat
- ! significant verschil in verwachte richting
- !! significant verschil in niet-verwachte richting
- ? verschil in verwachte richting, maar niet significant
- ?? verschil in niet-verwachte richting, maar niet significant

N.B. 'verwachte richting' houdt in: meer of minder in TO, al naar gelang de hypothese vermeldt.

N.B. De subcorpora R en StG zijn samen als corpus TO vergeleken met de corpora D, O of W uit *Woordfrequenties*.

ONPERSOONLIJK

<i>naamwoordstijl</i>	TO			
	R	StG	D	G
1. substantieven (+)	-	-	!	!
2. substantivisch gebruikte adjectieven (+)	-	-	??	??
3. substantivisch gebruikte infinitieven (+)	-	-	!	!
4. substantivisch gebruikte deelwoorden (+)	-	-	?	?
5. werkwoorden (-)	-	-	!	!
6. voorzetsels (+)	+	-	!	!
7. substantieven op '-ing' of '-ingen' (+)	-	-	!	!

lijdende vorm

8. lijdende vorm (+)	-	-	!	!
----------------------	---	---	---	---

het menselijk element

9. personalia (-)	-	+	!	!
10. possessiva (-)	-	-	!	!
11. verwijzingen naar een eerste persoon (-)	-	-	!	!
12. 'ondergetekende(n)' (+)	+	-	!	!
13. 'men' (+)	-	-	!!	!!
14. 'u' (-)	-	+	??	?
15. vraagzinnen (-)	-	+	!!	!!
16. eigennamen (-)	-	-	!	!
17. interjecties (-)	-	-	?	!
18. imperatieven (-)	-	-	?	!

INGEWIKKELD

bijzinnen en tussenzinnen

19. persoonsvormen (+)	-	+	!!	!!
20. zinslengte (+)	-	-	!	!
21. relativa (+)	-	-	!!	!!
22. betrekkelijke (voornaamwoordelijke) bijwoorden (+)	-	-	??	??
23. onderschikkende voegwoorden (+)	-	+	!	?
24. 'te' 650 (+)	-	-	?	?
25. nevenschikkende voegwoorden (+)	-	-	!!	!!
26. voegwoorden van vergelijking (+)	-	-	??	??

	TO				
	R	StG	D	O	W
<i>tangeconstructies</i>					
27. afstand tussen delen van voornaamwoordelijke bijwoorden (+)	-	-	!!	!!	??
28. afstand tussen delen van samengestelde werkwoorden (+)	-	-	??	??	??
29. afstand binnen voorzetselgroepen (+)	-	-	??	??	??
30. voegwoord plus voegwoord (+)	-	-	??	??	?
31. klemconstructies met deelwoord of infinitief (+)	-	-	!	!	!
32. aantal woorden per klemconstructie (+)	-	-	!	?	!
33. ketens van voorzetsels en lidwoorden (+)	-	-	!	!	!
<i>andere zinsbouwverschijnselen</i>					
34. ontkennende woorden (+)	-	-	??	!!	!!
35. woorden met 'niet-', 'non-' of 'on-' (+)	-	-	??	??	!!
<i>moeilijke woorden</i>					
36. woordlengte (+)	-	-	!	!	!
WIJDLOPIG					
37. adjectieven (+)	-	-	??	!!	!!
38. bijwoorden (+)	-	-	!!	!!	!!
TRADITIONEEL					
39. aanvoegende wijs (+)	-	-	!	!	?
40. substantieven in verouderde naamvals-vormen (+)	-	-	!	!	!
41. pronomina in verouderde naamvals-vormen (+)	-	-	?	?	!!
42. verbogen voorzetsels (+)	+	-	!	!	!
43. enkele archaïsmen (+)	-	-	!	!	!
<u>'Andere Uitkomsten'</u>					
13 ¹ . 'men'/indefinita, zelfstandig (+)	-	-	??	??	!!
14 ¹ . 'u'/personalia, reflexiva (-)	-	+	!!	!!	!!
18 ¹ . imperatieven/p'v's (-)	-	-	??	?	?
42 ¹ . verbogen voorzetsels/voorzetsels (+)	-	-	!	!	!

Op deze pagina volgt voor alle duidelijkheid nog een overzicht van hypothesen met dezelfde toetsingsresultaten. (In dit overzicht zijn **niet** de hypothesen opgenomen waarbij toetsing een afwijking in verwachte **en** in niet-verwachte richting te zien gaf, te weten 14, 18¹, 30 en 41.)

D	O	W
!	!	!
1. substantieven (+)		
6. voorzetsels (+)		
7. substantieven op '-ing' of '-ingen' (+)		
8. lijdende vorm (+)		
9. personalia (-)		
10. possessiva (-)		
11. verwijzingen naar een eerste persoon (-)		
12. 'ondergetekende(n)' (+)		
20. zinslengte (+)		
31. klemconstructies met deelwoord of infinitief (+)		
33. ketens van voorzetsels en lidwoorden (+)		
36. woordlengte (+)		
40. substantieven in verouderde naamvals-vormen (+)		
42. verbogen voorzetsels (+)		
43. enkele archaïsmen (+)		
!	!(?)	!(?)
3. substantivisch gebruikte infinitieven (+)		
5. werkwoorden (-)		
16. eigennamen (-)		
32. aantal woorden per klemconstructie (+)		
39. aanvoegende wijs (+)		
!(?)	!(?)	?
17. interjecties (-)		
18. imperatieven (-)		
23. onderschikkende voegwoorden (+)		
?	?	?
4. substantivisch gebruikte deelwoorden (+)		
24. 'te' 650 (+)		

D	O	W
!!	!!	!!
13. 'men' (+)		
14! 'u'/personalia, reflexiva (-)		
15. vraagzinnen (-)		
19. persoonsvormen (+)		
21. relativa (+)		
25. nevenschikkende voegwoorden (+)		
38. bijwoorden (+)		
??	??	!!
2. substantivisch gebruikte adjectieven (+)		
13! 'men'/indefinita, zelfstandig (+)		
26. voegwoorden van vergelijking (+)		
35. woorden met 'niet-', 'non-' of 'on-' (+)		
!!(??)	!!	!!(??)
27. afstand tussen delen van voornaamwoordelijke bijwoorden (+)		
34. ontkennende woorden (+)		
37. adjectieven (+)		
??	??	??
22. betrekkelijke (voornaamwoordelijke) bijwoorden (+)		
28. afstand tussen delen van samengestelde werkwoorden (+)		
29. afstand binnen voorzetselgroepen (+)		

7 DISCUSSIE

7.1 Inleiding

In deze inleiding staan opmerkingen over de operationalisering van de oordelen (7.1.1), de indeling van de hypothesen (7.1.2) en over zinslengte (7.1.3). Dit laatste punt is van belang omdat, door verschil in steekproeftrekking voor *Woordfrequenties* en dit onderzoek, de uitkomsten niet zonder meer met elkaar vergeleken kunnen worden. In 7.2 volgt de bespreking van de uitkomsten per hypothese of per aantal samenhangende hypothesen. In 7.3, Conclusies, staat een overzicht van de belangrijkste verschillen, waaronder ook verschillen tussen 'regering' en 'Staten-Generaal'. Hierbij zal vooral aandacht worden geschonken aan uitkomsten die in strijd zijn met de in hoofdstuk 3 vermelde intuïtieve oordelen. Deze studie besluit, in 7.4, met suggesties voor verder onderzoek.

7.1.1 De operationalisering van de oordelen

In hoofdstuk 6 is steeds nauwkeurig aangegeven wat er naar aanleiding van de oordelen over de taal van 'Den Haag' is onderzocht. Met nadruk zij hier nog vermeld dat binnen het gekozen beschrijvingsmodel, het codeersysteem C₃C, niet elk oordeel volledig en nauwkeurig geherformuleerd kon worden tot een hypothese. Vier voorbeelden:

1. Van de in hoofdstuk 3 genoemde tangconstructies zijn er enkele buiten beschouwing gelaten, omdat via het codeersysteem hierover geen informatie kon worden verkregen (o.a. de afstand tussen onderwerp en werkwoord, en tussen werkwoord en lijdend voorwerp).
2. Onder het trefwoord 'wijdlopig' is op pagina 41 het oordeel gegeven dat het bij pleonastisch taalgebruik vooral gaat om het overbodig gebruik van bijvoeglijke naamwoorden en bijwoorden. Maar met ons codeersysteem kan pleonastisch taalgebruik niet worden onderzocht. Wel kan gemakkelijk worden nagegaan of in TO meer adjectieven en bijwoorden voorkomen dan in D, O of W. Dit is dan ook gebeurd via de hypothesen 37 en 38. Maar op basis van het meer of minder voorkomen van deze categorieën woorden is geen uitspraak te doen over pleonastisch taalgebruik.
3. Bij de woorden op '-ing/-ingen', hypothese 7, zijn ook de woorden meegeteld die niet direct in verband gebracht kunnen worden met een werkwoord, bijvoorbeeld "Op dit terrein staan alleen onbewoonbaar ver-

klaarde woningen." Het is aannemelijk dat de te onderzoeken teksten op dit punt niet zullen verschillen. Maar hieruit mag niet geconcludeerd worden, dat de vergelijking zuiver is. Immers wanneer we twee teksten, A en B, hebben met in A 62 en in B 70 woorden op '-ing/-ingen', en wanneer neer elke tekst 60 woorden bevat, die niet direct in verband gebracht kunnen worden met een werkwoord (het type 'woning'), dan krijgen we een heel andere verhouding, namelijk A 2 en B 10. Een nauwkeuriger beschrijving is hier niet te geven, omdat het niet mogelijk is alle woorden van het type 'woning' eruit te halen. Dit is een goed voorbeeld van een bezwaar dat *Ullmann* (1964:118,119) noemt tegen de kwantitatieve stilistiek (zie in deze studie pagina 20 en 21), namelijk dat verschillende verschijnselen die oppervlakkig gezien gelijk zijn, onder een noemer worden gebracht. Hoe het ook zij, wat de woorden op '-ing/-ingen' betreft komen we niet verder. In elk geval moet worden vermeld dat deze categorie woorden, zoals geteld met behulp van de retrogradelijsten (dus zonder context), ruimer is dan de woorden op '-ing/-ingen' zoals bedoeld in hoofdstuk 3 bij de oordelen over '-ing-ziekte'. Zie verder de bespreking van deze hypothese op pagina 148.

4. Onder imperatieven, hypothese 18, vallen ook zinswendingen als "zie pagina 13". Het is heel goed mogelijk dat T0 en W op dit punt niet verschillen, omdat er in W meer van dit type verwijzingen voorkomen. Wanneer men een term als 'imperatieven' leest, zal men niet zo gauw geneigd zijn ook verwijzingen daaronder te vatten. Soms wordt zo iets pas duidelijk na lezing van de toelichting op het codeersysteem, hoofdstuk 4, of de opmerkingen bij de desbetreffende hypothese in hoofdstuk 6.

Bij elke hypothese moet bedacht worden dat die is opgesteld naar aanleiding van oordelen. In dit 'naar aanleiding van' zit een zekere vaagheid. Soms dekt de hypothese het oordeel vrij aardig, vooral als het oordeel specifiek is; soms kwam Procrustes er aan te pas om het oordeel in het beschrijvingsmodel te passen. In een enkel geval kon een beperking van het C₃C-systeem ongedaan gemaakt worden, bijvoorbeeld bij hypothese 40 over substantieven in verouderde naamvals vormen waar woorden als 's middags' en 's zomers' buiten beschouwing zijn gelaten. In een ander geval is aanvullend handwerk verricht op basis van de gecodeerde teksten, bijvoorbeeld bij hypothese 8 over de lijdende vorm. Maar soms was ook dat niet mogelijk; zie de onder punt 1 genoemde tangconstructies.

Een ander punt dat hier nog aan de orde moet komen is, dat onderscheidingen binnen het codeersysteem C₃C soms aanleiding geven tot hypothesen die niet direct terug te voeren zijn op de oordelen. Enkele voorbeelden:

1. In hoofdstuk 3 wordt onder 'naamwoordstijl' niet gesproken over substantivisch gebruikte adjectieven (hypothese 2) en substantivisch gebruikte deelwoorden (hypothese 4). Maar omdat het voorkomen van deze categorieën ook aan het oordeel 'naamwoordstijl' ten grondslag kan liggen, zijn hierover aparte hypotheses opgesteld.
2. Onder 'het menselijk element' is niets vermeld over het minder voorkomen van eigennamen (hypothese 16), interjecties (hypothese 17) of imperatieven (hypothese 18). Maar omdat deze onderscheidingen in het codeersysteem aanwezig waren en omdat het aannemelijk is dat deze categorieën minder voorkomen wanneer 'het menselijk element' ontbreekt, zijn ook hierover hypotheses opgesteld.
3. Hetzelfde kan gezegd worden van enkele hypotheses genoemd onder 'ingewikkeld' en 'traditioneel', o.a. de hypotheses 27, 28, 29 en 40, 41, 42.

Samenvattend kan gesteld worden dat het C₃C-systeem voor het operationaliseren van oordelen over de taal van 'Den Haag' soms te weinig en soms een overdaad aan onderscheidingen geeft. Maar dit bezwaar is niet te ondervangen binnen de opzet van dit onderzoek. Van een algemeen beschrijvingsmodel kan niet worden verwacht dat het alle kenmerken van een speciaal type taalgebruik in kaart brengt. Het voordeel van een algemene beschrijving is, dat er gemakkelijk kan worden vergeleken, ook met nog te onderzoeken typen taalgebruik. Dit zou veel moeilijker zijn, wanneer er een specifiek codeersysteem was ontworpen om alleen de verschijnselen in kaart te brengen, die genoemd zijn in oordelen over de taal van 'Den Haag'.

7.1.2 De indeling van de hypotheses

In hoofdstuk 3 zijn de oordelen gerangschikt onder vier positieve en vier negatieve trefwoorden: (positief) zakelijk, bondig, duidelijk en waardig; (negatief) onpersoonlijk, ingewikkeld, wijldlopig en traditioneel. Deze indeling had als doel de grote verscheidenheid aan oordelen zo overzichtelijk mogelijk te presenteren. Maar in het vorige hoofdstuk, en straks ook bij de bespreking van de resultaten, keert deze indeling terug.

Al te gemakkelijk zou de conclusie getrokken kunnen worden, dat bijvoorbeeld de hogere frequentie van voorzetsels in T0, hypothese 6 onder naamwoordstijl, mede verantwoordelijk is voor het onpersoonlijke karakter van overheidsteksten, of dat door de niet-vermoede lagere frequentie van persoonsvormen, hypothese 19, T0 minder ingewikkeld is dan D, O of W.

Met dit type conclusies, het relateren van een taalverschijnsel aan een algemeen, in vage termen gegeven, oordeel over taalgebruik, moet men voorzichtig zijn. Op pagina 33 is al aangegeven dat sommige specifieke oordelen over zinsbouw en woordkeus in verband gebracht kunnen worden met verschillende algemene oordelen. Het is heel goed mogelijk dat de hoge frequentie van voorzetsels ook te maken heeft met het grote aantal klemconstructies, en zo leidt tot het algemene oordeel 'ingewikkeld taalgebruik'. Door de aard van de oordelen zijn op dit punt heel moeilijk nauwkeurige en verantwoorde uitspraken te doen. Nader onderzoek zou moeten uitwijzen of bijvoorbeeld de verschijnselen genoemd in de vijf laatste hypothesen, van 'aanvoegende wijs' (hypothese 39) tot 'enkele archaïsmen' (hypothese 43) werkelijk aanleiding geven tot het oordeel 'traditioneel'. De mogelijkheid is niet uitgesloten dat een tekst waarin deze verschijnselen niet voorkomen, toch als 'traditioneel' beoordeeld wordt op grond van bijvoorbeeld bepaalde andere kenmerken in woordkeus. En omgekeerd, een tekst met de archaïsmen uit hypothese 43 hoeft nog niet het oordeel 'traditioneel' te ontlokken, omdat deze archaïsmen gecompenseerd kunnen worden door andere 'moderne' woorden.

Maar over de indeling van de hypothesen valt nog meer op te merken. Bijvoorbeeld, hypothese 21 en 31 kunnen met elkaar in tegenspraak zijn. Wanneer er veel klemconstructies voorkomen, *het in deze nota weergegeven standpunt*, dan is het aannemelijk dat er minder relativa voorkomen, *het standpunt dat 420 in deze nota is weergegeven*, tenzij uiteraard naast veel bijvoeglijke bijzinnen ook nog veel klemconstructies gebruikt worden. Hetzelfde geldt voor hypothese 5 en 19, minder werkwoorden naast meer persoonsvormen. Dit zou alleen kunnen wanneer het in hypothese 5 om niet-persoonsvormen ging. Deze tegenstrijdigheden zijn terug te voeren op de oordelen. Op pagina 37 is al opgemerkt dat het oordeel 'ingewikkelde zinsbouw' te maken heeft met veel bepalingen enerzijds en met veel bij- en tussenzinnen (vandaar de hypothesen over relativa en persoonsvormen) anderzijds. In principe kunnen natuurlijk beide opmerkingen waar zijn. Daarom zijn er hypothesen opgesteld die op het eerste gezicht met elkaar in strijd lijken te zijn.

Een aantal andere hypothesen zijn onderling afhankelijk. In het vorige hoofdstuk (pagina 125) is al vermeld dat ketens van voorzetsels en lidwoorden, hypothese 33, voor het grootste gedeelte vallen onder de in hypothese 31 genoemde klemconstructies. Ook verbogen lidwoorden en verbogen voorzetsels zijn voor een deel afhankelijk van zelfstandige naamwoorden in archaïsche vorm, of omgekeerd: *de heer des huizes*, *ten huize van*. Maar ook bij deze eventuele onderlinge afhankelijkheid geldt,

dat het in eerste instantie beter is zoveel mogelijk gesignaleerde verschijnselen apart te onderzoeken, dan bij voorbaat mogelijke hypothesen uit te sluiten.

7.1.3 Zinslengte

In 5.1.1 is al opgemerkt dat de onderzoeksresultaten over zinslengte in het corpus TO niet direct te vergelijken zijn met gegevens die zijn af te leiden uit *Woordfrequenties*. Dit komt doordat de corpora D, O en W op een andere manier zijn samengesteld.

Voor TO zijn de fragmenten geselecteerd volgens de procedure beschreven in 5.2.3: vier zinnen meenemen, acht zinnen overslaan enz. Hierbij werden de afgebroken zinnen aan het einde of het begin van een geselecteerde pagina buiten beschouwing gelaten. Maar omdat een lange zin meer kans heeft om afgebroken te worden, bevat de steekproef meer korte zinnen dan de populatie. De gemiddelde zinslengte van de steekproef zal dus naar alle waarschijnlijkheid lager zijn dan die in de populatie.

Achteraf bleek bij controle dat het aantal afgebroken zinnen zo klein was, namelijk 26 in R en 8 in StG, dat gesteld mag worden dat de steekproef bijna representatief is (zie de tabel op de volgende pagina). Maar bij een volgend onderzoek is het beter, uitgaande van deze methode, de pagina ná de geselecteerde pagina te bewaren, om afgebroken zinnen die eigenlijk in de steekproef moeten worden opgenomen, alsnog te kunnen aanvullen.

Voor D, O en W is de instructie voor de selectie van de fragmenten voor meer dan één uitleg vatbaar. Uit het 'Verslag van de Werkgroep' (1974: 19-28) blijkt dat er drie mogelijkheden zijn:

1. Prik een regel. De zin waarin geprikt wordt, telt niet mee. Neem een fragment dat begint bij de volgende zin, met een omvang die zo dicht mogelijk komt bij 75 woorden (D), 115 woorden (O) of 250 woorden (W).
2. Als methode 1 voor D en O. Maar neem een fragment niet mee, wanneer het kleiner dan 60 of groter dan 90 is (bij D) of kleiner dan 90 en groter dan 140 (bij O).
3. Prik een regel. De zin waarin geprikt wordt, telt wel mee. Neem een fragment dat begint bij die zin, van minimaal 75 (D), 90 (O) of 250 (W) woorden; dat wil zeggen zoveel zinnen als nodig zijn om deze aantallen te bereiken.

Een lange zin loopt over meer regels dan een korte zin. Dus bij het prikken van een regel heeft een lange zin meer kans dan een korte. Met andere woorden: methode 1 levert in de steekproef kortere zinnen op dan de populatie te zien geeft; methode 2 levert nog kortere zinnen op,

want een fragment dat weggegooid wordt zou anders een heel lange zin bevatten (er zit geen zinseinde tussen 60 en 90 bij D, en tussen 90 en 140 bij O); methode 3 levert langere zinnen op. Of omgekeerd, bij methode 1 en 2 is de werkelijke zinslengte groter dan uit de steekproef blijkt en bij methode 3 juist kleiner. In dit verband zij er nog op gewezen dat zinslengte van invloed kan zijn op de woordsoortenverdeling. Bijvoorbeeld: hoe langer de zin, des te groter de kans dat er meer voegwoorden in voorkomen. Wanneer, door de selectiemethode, de steekproef qua zinslengte afwijkt van de populatie, worden dus ook de gegevens over woordsoortenverdeling minder betrouwbaar!

Helaas is op basis van het 'Verslag van de Werkgroep' niet precies na te gaan welke methode, of welke mengvorm van methoden, is gevolgd. Wel kan met behulp van formules ontwikkeld door *Gill* (te verschijnen), geschat worden hoe de populatie eruit ziet bij een gegeven steekproefverdeling, verkregen volgens een van deze methoden; dit op basis van de steekproefverdeling bij een gegeven populatie. In de volgende tabel staat aangegeven hoe de gemiddelde zinslengte in de steekproeven afwijkt van schattingen van de gemiddelde zinslengte in de populaties, uitgaande van de veronderstelling dat de steekproeven zijn verkregen volgens de verschillende selectiemethoden.

	D	O	W	T0	(R)	(StG)
1.	18,30	18,41	21,08	24,64	25,05	23,58
2.	-	-	-	<u>24,76</u>	<u>25,17</u>	<u>23,68</u>
3.	18,94	19,07	<u>21,44</u>			
4.	<u>19,13</u>	<u>19,17</u>				
5.	17,39	17,42	20,48			
1. gemiddelde zinslengte in de steekproeven 2. geschatte gemiddelde zinslengte in de populatie T0 3. geschatte gemiddelde zinslengte, uitgaande van selectiemethode 1 4. geschatte gemiddelde zinslengte, uitgaande van selectiemethode 2 5. geschatte gemiddelde zinslengte, uitgaande van selectiemethode 3						

De beste interpretatie van het 'Verslag van de Werkgroep' is de volgende. Bij D en O is er geselecteerd volgens methode 2, wanneer er in een zin werd geprikt. Bij W is in die gevallen methode 1 gevolgd. Toch zal bij D en O de waarheid vermoedelijk liggen tussen de uitkomsten van methode 2 en de gegevens uit de steekproef, omdat vaak tussen tekstgedeelten

(in foto's, koppen of regels wit) wordt geprikt, dat wil zeggen niet in een zin. Bij W ligt de waarheid dichterbij methode 1 omdat hier minder sprake zal zijn van prikken tussen tekstgedeelten. Dus het verschil in gemiddelde zinslengte tussen bijvoorbeeld TO en D is ten minste 5,63 woorden ($24,76 - 19,13$) en ten hoogste 6,46 woorden ($24,76 - 18,30$). Bij het toetsen van de hypothese over zinslengte zijn dan ook de in de tabel onderstreepte getallen gehanteerd.

In Bijlage 4 (pagina 220 en 221) vindt u de grafische voorstelling van de verschillen in zinslengte tussen de steekproeven uit D, O en W en de vermoede zinslengte in de populaties D, O en W. In de grafieken is weer-gegeven de zogenaamde empirische cumulatieve verdelingsfunctie. Dit houdt in dat met behulp van deze grafieken antwoord gegeven kan worden op de vraag: Hoe groot is de kans dat men een zin prikt van een bepaalde lengte of een zin die korter is dan die lengte? Op de y-as staat de kans en op de x-as de zinslengte. Een voorbeeld: Hoe groot is de kans dat men in D een zin prikt van 15 of minder woorden? Voor D is de geschatte kans voor de populatie, uitgaande van de veronderstelling dat er geselecteerd is volgens methode 2, ongeveer 42 procent. Voor de steekproef D is deze kans 44 procent. (Hoe meer naar rechts de lijn ligt, des te langer zijn de zinnen.)

In de laatste grafiek van Bijlage 4 staan de zinslengten van de verschillende populaties bij elkaar. (TO liggend tussen R en StG, is hier niet opgenomen.) Wanneer we nu de vraag willen beantwoorden hoe groot de kans is dat men een zin prikt van 24 of minder woorden, vinden we de volgende kansen: R, 54 procent; StG, 59 procent; W, 66 procent; O, 73 procent en D, 74 procent. Uit deze laatste grafiek, waarin D en O elkaar kruisen, valt ook af te leiden, dat de kans op een zin van bijvoorbeeld 12 woorden of minder in O hoger is dan in D, terwijl de kans op een zin van bijvoorbeeld 26 woorden of minder in O lager is dan in D.

7.2 Bespreking van de uitkomsten

7.2.1 Onpersoonlijk

7.2.1.1 Naamwoordstijl

Over dit verschijnsel zijn naast de hypothese over substantieven, drie andere opgesteld die ook met naamwoordstijl in verband gebracht kunnen worden, namelijk de hypothesen 2, 3 en 4 over substantivisch gebruikte adjectieven, infinitieven en deelwoorden.

De hypothese over de substantivisch gebruikte adjectieven, code 12., *de goeden 121, de jongeren 126*, wordt niet bevestigd. Integendeel, deze categorie komt in W significant meer voor dan in TO. Bij deze hypothese moet nog het volgende worden opgemerkt. Het is zeer de vraag of deze categorie woorden wel te maken heeft met naamwoordstijl (zie ook pagina 143). Want bij naamwoordstijl gaat het vooral om het gebruik van zelfstandige naamwoorden in plaats van werkwoorden. Dit is bij code 12. eigenlijk nooit het geval. (Onder 12. vallen bijvoorbeeld namen voor talen en kleuren.) In een eerder stadium van het onderzoek was deze categorie woorden opgeteld bij hypothese 1 over substantieven. Ook toen gaf hypothese 1 hetzelfde resultaat als nu (zie beneden). De extra informatie die we nu hebben verkregen door een aparte hypothese is, dat substantivisch gebruikte adjectieven in TO minder voorkomen dan in W. Bij de substantivisch gebruikte infinitieven, *het toetsen 201 van een hypothese*, blijkt het volgende. In TO komen significant meer substantivisch gebruikte infinitieven voor dan in D en O. Tussen TO en W is er geen significant verschil; wel valt tussen de corpora TO en W een verschil in verwachte richting te constateren. Bij de substantivisch gebruikte deelwoorden, *het volgende 203, de verdachte 217*, is er geen significant verschil.

Tot zover de drie hypothesen naast hypothese 1 over de substantieven. Het is overigens opvallend dat de categorieën woorden uit de twee hypothesen die niet worden bevestigd, substantivisch gebruikte adjectieven en deelwoorden, niet worden genoemd in de oordelen in hoofdstuk 3; dit in tegenstelling tot de substantivisch gebruikte infinitieven (zie hierover pagina 34).

Hypothese 1 over substantieven wordt bevestigd. In deze hypothese zijn ook opgenomen de woorden met code 021, maar deze adjectivisch gebruikte substantieven, *het minimum 020 inkomen*, komen bijzonder weinig voor (zie Bijlage 2) en oefenen dus amper invloed uit.

Dat hypothese 6, meer voorzetsels, wordt bevestigd zal na de uitleg van hypothese 1 geen verwondering wekken. Voorzetsels dienen immers (*Den Hertog*, 1973, dl. III:232) "om een substantief of substantief gebruikt woord als bijvoeglijke of bijwoordelijke bepaling, of als omschreven voorwerp te gebruiken."

Het zou bijzonder interessant zijn om na te gaan of het verschil in substantieven wordt veroorzaakt door bepaalde typen substantieven. In elk geval is er een significant verschil bij de woorden op '-ing/-ingen', hypothese 7, niet alleen vergeleken met het totaal, maar ook binnen de categorie substantieven met code 000 en 001. Hierbij moet wel worden

opgemerkt (zie paragraaf 7.1.1) dat lang niet alle substantieven op '-ing/-ingen' in verband gebracht kunnen worden met de naamwoordstijl. Dit is alleen het geval wanneer er in plaats van zo'n substantief een werkwoord had kunnen staan. Bijvoorbeeld wel in "zij komen tot de constatering dat ..." (zij constateren dat ...) maar niet in "een onbewoonbaar verklaarde woning". Eigenlijk zou bij elk substantief van dit type de context moeten worden bekeken om na te gaan of er sprake is van een woord dat vervangen kan worden door een werkwoord. Maar omdat beslissingen hierover in hoge mate arbitrair zijn, heb ik een poging hiertoe al spoedig gestaakt.

Het vermoeden dat het verschil in substantieven vooral wordt veroorzaakt door gesubstantiveerde werkwoorden, vindt steun in het toetsingsresultaat van hypothese 5, werkwoorden (zoals omschreven in de hypothese, dus geen verbogen deelwoorden, hulpwerkwoorden enz.). TO bevat significant minder werkwoorden dan D en O. Tussen TO en W is er echter geen significant verschil.

Nader onderzocht is nog, of er ook verschil zit in de categorie hulp- en koppelwerkwoorden. (Deze categorie is niet in hypothese 5 opgenomen.) In de naamwoordstijl zou deze volgens de oordelen uit hoofdstuk 3 meer moeten voorkomen. Maar hier bleek geen significant verschil.

7.2.1.2 Lijdende vorm

Ook hier een significant verschil: In TO komt de lijdende vorm meer voor dan in D, O of W.

Bij de getallen die gegeven zijn voor de lijdende vorm, moet wel een zekere marge in acht worden genomen, gezien het soms moeilijke onderscheid tussen 'zijn' als koppelwerkwoord en 'zijn' als hulpwerkwoord van de lijdende vorm. Bijvoorbeeld *Hij was 275 al jarenlang verlamd* (koppelwerkwoord) en *Hij was 275 verlamd door de schrik* (hulpwerkwoord van de lijdende vorm). Maar omdat het percentage twijfelgevallen per categorie onderzochte teksten niet verschilt, geven de getallen relatief wel een zuiver beeld.

Verder moet nog opgemerkt worden dat voor D, O en W alleen de eerste 1955 zinnen zijn onderzocht. (TO bevat ook 1955 zinnen.) Dit houdt in dat de zinnen uit de corpora D, O en W niet aselekt zijn gekozen.

Maar we mogen, gezien de redelijke homogeniteit in D, O en W, veilig aannemen dat dit eerste gedeelte (ongeveer een derde deel) representatief is. Een vergelijking op dit punt geeft dus geen problemen.

Op een ander punt leverde de vergelijking wel moeilijkheden op. Waar-
mee moeten de lijdende vormen, of beter gezegd de woorden die wijzen op
een lijdende vorm, worden vergeleken? Het aantal lijdende vormen kan
afgezet worden tegen het aantal zinnen, wanneer per zin bekend is hoe-
veel lijdende vormen erin voorkomen. Maar een eventueel verschil in
zinslengte tussen T0 enerzijds en D, O of W anderzijds kan dan ook de
oorzaak zijn voor meer lijdende vormen in T0. (Hoe langer de zin, des
te groter de kans op meer hoofd- en bijzinnen, die mogelijkerwijs in de
lijdende vorm staan.)

Een andere mogelijkheid is: Woorden die wijzen op een lijdende vorm,
vergelijken met woorden die veranderen wanneer de zin lijdend wordt ge-
maakt: de persoonsvormen. Maar ook hier doen zich problemen voor. Immers
de categorie woorden die verwijst naar een lijdende vorm, bevat niet
alleen persoonsvormen maar ook infinitieven (code 230) van 'worden'
en 'zijn'. Vergelijk *De regering zal 273 maatregelen nemen* en *Door de*
regering zullen 274 maatregelen worden 230 genomen. Wanneer gekozen
wordt voor vergelijking met persoonsvormen, houdt dit in dat een cate-
gorie woorden waarin ook niet-persoonsvormen zitten, namelijk de code
230, vergeleken wordt met een categorie waarin alleen persoonsvormen
zitten. Het lijkt er dan, populair gezegd, op, of hier het getal bepaald
wordt van een bepaald soort appels samen met peren, binnen een groter ge-
heel van appels zonder peren. Bovendien kunnen woorden met code 26.,
reflexief gebruikte persoonsvormen, nooit veranderd worden in een con-
structie met een lijdende vorm, omdat volgens het Codeervoorschrift
transitief altijd gaat boven reflexief (zie pagina 55). Ook een gedeelte
van code 27. namelijk de koppelwerkwoorden, moet dan buiten beschouwing
blijven om dezelfde reden: *Hij is 273 ziek* kan nooit veranderd worden
in een zin met een lijdende vorm. Code 26. kan gemakkelijk buiten-
gesloten worden, maar het zal duidelijk zijn dat dit bij het zojuist
genoemde gedeelte van code 27. veel moeilijker is.

Gezien de hier vermelde problemen bij het vergelijken, zijn de aantal-
len lijdende vormen afgezet tegen het aantal woorden, dat wil zeggen
voor D, O en W een even groot percentage woorden als er zinnen genomen
zijn, ongeveer een derde deel.

Uit nader onderzoek bleek ook, dat het oordeel over de lijdende vorm
niet getoetst mag worden op basis van 'worden' en 'zijn' als hulp- of
koppelwerkwoord (zie pagina 109). Dit vanuit de veronderstelling dat D,
O, W en T0 niet zullen verschillen in het voorkomen van 'worden' en

'zijn' als koppelwerkwoord en als hulpwerkwoord van tijd, en dat eventuele verschillen dus kunnen worden toegeschreven aan het voorkomen van 'worden' en 'zijn' als hulpwerkwoord van de lijdende vorm. Want wanneer we het percentage berekenen van 'worden' en 'zijn' als hulp- en koppelwerkwoord (door de getallen in rij 2 van de tabel op pagina 109 te delen door het totaal aantal woorden) en hetzelfde doen voor 'worden' en 'zijn' alleen als hulpwerkwoord van de lijdende vorm (op basis van het aanvullend handwerk zoals beschreven op pagina 110 en 111), en dan dit laatste percentage van het eerste aftrekken, dan blijkt dat er alle reden is om te veronderstellen dat TO minder vormen van 'worden' en 'zijn' als koppelwerkwoord en als hulpwerkwoord van tijd bevat dan D, O en W. Met andere woorden: door het aanvullend handwerk krijgen we een zuiverder beeld.

Tot slot nog het volgende. Bij het onderzoek naar de lijdende vorm is ook geteld hoe vaak er sprake was van een passieve constructie met een intransitief voltooid deelwoord, *Er wordt 273 gedacht 206 aan*, hoe vaak de 'door-bepaling' werd gebruikt en hoe vaak er een of twee infinitieven optraden, bijvoorbeeld *Hierbij kan 273 worden 230 opgemerkt; Hierbij zou 275 opgemerkt kunnen 230 worden 230*. Hieronder ter nadere informatie de percentages uit de corpora.

	D	O	W	TO	(R)	(StG)
1.	3,8	5,2	2,8	8,3	9,1	5,8
2.	18,9	15,4	15,0	11,3	10,2	14,7
3.	20,4	17,7	20,7	31,0	29,9	34,7
4.	1,5	1,8	1,4	4,6	4,4	5,4

1. lijdende vorm met intransitief gebruikt voltooid deelwoord
2. lijdende vorm met 'door-bepaling'
3. lijdende vorm met één infinitief
4. lijdende vorm met twee infinitieven

In corpus TO komen dus meer lijdende vormen met een intransitief werkwoord voor en minder 'door-bepalingen'. Bovendien lijken de lijdende vormen ingewikkelder doordat er meer infinitieven bij worden gebruikt.

7.2.1.3 Het menselijk element

De hypothesen over personalia en possessiva die verwijzen naar personen worden beide bevestigd: In TO komen deze woorden minder voor. Nader onderzoek wees uit dat ook de volledige categorieën personalia en possessiva (dus met inbegrip van verwijzingen naar zaken) in TO significant minder voorkomen.

Het minder voorkomen van personalia komt overeen met de uitslag van hypothese 19, minder persoonsvormen. En in plaats van possessiva worden waarschijnlijk meer demonstrativa gebruikt. (Zie Bijlage 2 bij code 370.) Hypothese 11, verwijzingen naar een eerste persoon, wordt eveneens bevestigd, ook wanneer deze woorden worden vergeleken met de woordsoorten waartoe ze behoren: personalia, possessiva en reflexiva. Dit wordt niet veroorzaakt door het feit dat in TO andere woorden worden gebruikt om te verwijzen naar een eerste persoon; zie hypothese 12 over *ondergetekende(n)*. Want wanneer woorden als *ondergetekende(n)*, *onzerzijds* enz. worden opgeteld bij de aantallen in hypothese 11, blijft de uitslag van de toetsing voor D, O en W gelijk. Wel blijkt dan een significant verschil tussen R en StG: In R komen de woorden genoemd in hypothese 11 en 12, samen meer voor dan in StG. Dit is ook niet verwonderlijk, omdat het woord *ondergetekende(n)* alleen in subcorpus R voorkomt (met een frequentie van 69).

Opvallend is, dat hypothese 13 over *men* een significante afwijking in niet-verwachte richting geeft. Wanneer *men* alleen vergeleken wordt met zelfstandige indefinita (hypothese 13¹) verschilt TO niet significant van D en O. *Men* geldt in studies voor leesbaarheid en bij adviezen over 'persoonlijk schrijven' als een woord dat typerend is voor een onpersoonlijke stijl. Maar dit woord komt in TO minder vaak voor dan in D, O of W. Het oordeel dat TO een onpersoonlijke stijl heeft kan dus niet gerelateerd worden aan een hoge frequentie voor *men*. Hieruit blijkt wel dat het de moeite waard is om oordelen, hoe vaak ook vermeld in literatuur, te controleren!

Bij *u* is het verschil tussen vergelijking met het totaal aantal woorden (hypothese 14) en alleen de personalia en reflexiva (hypothese 14¹) heel groot. Toetsing van 14¹ levert een significante afwijking in niet-verwachte richting op voor D, O en W. Maar bij 14 is er geen significant verschil. Wel komt in StG *u* significant meer voor dan in R. Maar hoe dan ook, het oordeel dat TO een onpersoonlijke stijl heeft door de lage frequentie van *u* wordt niet bevestigd.

Ook hypothese 15, meer vraagzinnen, wordt niet bevestigd, sterker nog:

Er is een significant verschil in niet-verwachte richting. Maar dit kan gemakkelijk worden verklaard. In StG zitten veel kamervragen; ruim een kwart van de zinnen in subcorpus StG is vragend. Dit heeft uiteraard grote invloed op het aantal vraagzinnen in corpus TO. In subcorpus R daarentegen is slechts 1 van de 1416 zinnen vragend. Het betreft hier een vraag met retorisch karakter: "Betekent dit echter niet, dat al wel gekozen is voor een einddoel - al dan niet bewust - dat in deze nota niet is geëxpliciteerd?" Wanneer we alleen R vergelijken met D, O en W, wordt de hypothese bevestigd voor O en W (met een afwijking in verwachte richting voor corpus D).

Hypothese 16, minder eigennamen, wordt bevestigd voor D en O. W staat op dit punt dicht bij TO. Hypothese 17, minder interjecties, wordt merkwaardigerwijs wel bevestigd voor O, maar niet voor D en W. Het ontbreken van significantie valt hier te verklaren, wanneer we ons realiseren dat ook bijvoorbeeld het antwoord 'Ja' op een kamervraag, code 800 krijgt. Overigens komt in corpus TO van de codes die te maken hebben met interjecties alleen code 800 voor. Dat O hier hoog scoort, is niet te verwonderen gezien het naar verhouding hoge aantal uitroepen, dat bij het onderzoek naar vraagzinnen aan het licht kwam (zie pagina 116).

Bij hypothese 18, minder imperatieven, is er verschil tussen vergelijking met het totaal aantal woorden (18) en met alleen persoonsvormen (18¹). Bij 18 is er alleen significant verschil tussen TO en O, maar bij 18¹ ontbreekt significantie. Wel geeft corpus O een noemenswaardige afwijking in verwachte richting. De toetsingsresultaten kunnen hier voor W wellicht verklaard worden uit het feit dat ook literatuur- en paginaverwijzingen als imperatieven worden gecodeerd.

7.2.2 Ingewikkeld

7.2.2.1 Bijzinnen en tussenzinnen

Opvallend is dat slechts een van de acht hypotheses voor D, O en W bevestigd wordt, namelijk hypothese 20: In TO zijn de zinnen over het algemeen langer. De gemiddelde zinslengte in TO komt overigens overeen met het resultaat van de globale verkenning voorafgaand aan de steekproeftrekking, namelijk 25 woorden (zie pagina 88).

Bij hypothese 23, onderschikkende voegwoorden, is er alleen significant verschil tussen TO en D. Wel bleek uit nader onderzoek naar aanleiding van (zie pagina 37) "het nimmer ophoudende woordje dat" dat *dat* als

onderschikkend voegwoord in TO significant meer voorkomt dan in D en W; voor corpus O geldt hier alleen verschil in verwachte richting. Ook bij hypothese 24 over *te* met code 650, is er geen significant verschil. De zinnen zijn dus niet langer door meer onderschikkende voegwoorden (met uitzondering van D) of constructies met *te*+infinitief, die veel in beknopte bijzinnen voorkomen.

Ook de resultaten van de andere hypothesen helpen ons niet verder bij het verklaren van het verschil in zinslengte. Integendeel, hier is steeds verschil in niet-verwachte richting, en in de meeste gevallen is dit verschil nog significant ook: TO bevat minder persoonsvormen, minder relativa en minder nevenschikkende voegwoorden. Bij de voegwoorden van vergelijking is er ook verschil in niet-verwachte richting, maar alleen bij W is dit verschil significant. En bij de betrekkelijke (voornaamwoordelijke) bijwoorden is er geen significant verschil. Ook *dat* als relativum is nog apart onderzocht. Hier bleek eveneens geen significant verschil.

Samenvattend: In TO zijn de zinnen over het algemeen langer dan in D, O of W, maar dit betekent niet dat de zinsstructuur in termen van hoofd- of bijzinnen ingewikkelder is (met een uitzondering ten opzichte van D voor de onderschikkende voegwoorden). Integendeel, er is reden om aan te nemen dat die structuur eenvoudiger is. Wel mag geconcludeerd worden dat de hoofd-, bij- en tussenzinnen in TO langer zijn. Dit wordt ook ondersteund door de uitkomsten van hypothese 1 en 6: 'meer substantieven' en 'meer voorzetsels' maken dat er meer bepalingen in een zin staan.

7.2.2.2 Tangconstructies

Naar aanleiding van de oordelen dat in teksten van de overheid zinsdelen die bij elkaar horen, vaak ver van elkaar staan, zijn drie hypothesen opgesteld over verschijnselen die niet met name worden genoemd in hoofdstuk 3, Oordelen over de taal van 'Den Haag'. Het betreft hier de hypothesen 27, 28 en 29 over de afstand tussen delen van voornaamwoordelijke bijwoorden, delen van samengestelde werkwoorden en delen van voorzetselgroepen. Deze hypothesen zijn niet alleen opgesteld omdat het codeersysteem hiertoe de mogelijkheid bood (de codes 610, 620 en 630), maar ook omdat men, op basis van het algemene oordeel over het uiteenplaatsen van zinsdelen, mag verwachten dat dit uiteenplaatsen ook voorkomt bij de hier genoemde categorieën woorden. Voor een goed begrip volgt hier nog een zin waarin de codes 610, 620 en 630 zijn verwerkt: *Het energieprobleem riep 2450 tot 600 nu toe 630 niet alleen op 6205 tot bezinning - daar 540 waren we al lang mee 6105 bezig -*

maar het verplicht ons ook nieuwe plannen uit 6200 te denken 2105.

(Zie voor het vierde cijfer bij de delen van voornaamwoordelijke bijwoorden en samengestelde werkwoorden de uitleg op pagina 68.)

Met de hypothesen 27, 28 en 29 is bijvoorbeeld verondersteld dat in TO meer woorden staan tussen 'tot' en 'toe', dat in TO 'op' eerder achter 'bezinning' staat en dat in D, O of W 'mee' dichterbij 'daar' staat. Deze hypothesen worden niet bevestigd. Sterker nog: Steeds blijkt een afwijking in niet-verwachte richting. Bij hypothese 27 over de afstand tussen delen van voornaamwoordelijke bijwoorden was deze afwijking voor D en O zelfs significant. Verder kan nog opgemerkt worden dat code 630 in corpus TO relatief veel minder voorkomt.

Ook de hypothese over een tangconstructie die ontstaat doordat een bijzin in een hoofd- of bijzin wordt opgenomen, hypothese 30 'voegwoord plus voegwoord', wordt niet bevestigd. Deze uitslag komt overeen met de resultaten van de hypothesen 21, 22 en 23 over bijzinnen.

Hypothese 31 over klemconstructies met deelwoord of infinitief wordt wel bevestigd. Dus in TO meer constructies van het type "de in dit hoofdstuk getrokken conclusies". Dit klopt ook met de uitslag van de hypothesen 19 en 21: minder persoonsvormen en minder relativa. In D, O en W zullen we vaker een constructie aantreffen als "de conclusies die in dit hoofdstuk zijn getrokken". Hierbij zij nog opgemerkt dat een klemconstructie meestal korter is dan een overeenkomstige bijzin, maar dat desondanks in TO de zinnen over het algemeen langer zijn. Bij het tellen van de klemconstructies kon ook gemakkelijk onderzocht worden hoe vaak er constructies voorkomen van het type "de te nemen maatregelen". Het oordeel dat in ambtelijke taal deze constructie 'algemeen gebruikelijk' is (zie pagina 39) is ondergebracht bij de hypothese over klemconstructies. Het betreft hier een klemconstructie van één woord, en wel met 'te'. Hier volgen ter nadere informatie de frequenties van deze constructie in de verschillende corpora: D, 1; O, 2; W, 9; TO, 21; R, 19; StG, 2. Dus constructies van het type "de te nemen maatregelen" komen in corpus TO aanzienlijk meer voor.

Nauw verband met hypothese 31 houdt hypothese 33 over ketens van voorzetsels en lidwoorden. Op pagina 125 is er al op gewezen dat de meeste van deze constructies voorkomen in klemconstructies. Het is dan ook niet verwonderlijk dat deze hypothese wordt bevestigd. Dat de klemconstructies in TO langer zijn, zien we bevestigd in de uitslag van hypothese 32, althans voor D en W. Wat corpus O betreft, is er alleen afwijking in verwachte richting.

7.2.2.3 Andere zinsbouwverschijnselen

Hier zijn alleen hypothesen opgesteld over ontkennende woorden (hypothese 34) en woorden met 'niet-', 'non-' of 'on-' (hypothese 35). Deze hypothesen worden niet bevestigd. In TO komen zelfs significant minder ontkennende woorden voor dan in O en W. Bij de woorden met 'niet-', 'non-' en 'on-' is er significant verschil in niet-verwachte richting voor W.

7.2.2.4 Moeilijke woorden

De hypothese over woordlengte wordt wel bevestigd. Hoewel de verschillen klein zijn, een gemiddelde woordlengte van 5,57 in corpus TO tegenover ongeveer 5,3 in de corpora D, O en W, zijn ze toch niet aan het toeval te wijten.

7.2.3 Wijdlopij

Opvallend is dat de hypothesen over adjectieven en bijwoorden niet bevestigd worden. TO bevat significant minder van deze woorden dan D, O of W (met uitzondering van D voor de adjectieven). Uiteraard is hiermee nog niets gezegd over pleonastisch taalgebruik (zie 7.1.1). Wel zouden we op basis van de uitkomsten van hypothese 1, 'meer substantieven' ook meer adjectieven mogen verwachten.

In dit verband wil ik nog wijzen op de kenmerken die Bernstein geeft voor de zogenaamde 'restricted code' en de 'elaborated code' (zie Appel e.a., 1979:147-162). De 'restricted code' die gehanteerd wordt door kinderen met een taalachterstand - althans volgens de aanhangers van de deficiet-opvatting - wordt, in tegenstelling tot de 'elaborated code', volgens Bernstein, onder andere gekenmerkt door een rigide en beperkt gebruik van adjectieven en bijwoorden. Vermoedelijk wordt hier niet alleen het geringe aantal adjectieven en bijwoorden bedoeld, maar ook het geringe aantal verschillende woorden binnen deze categorieën. Gezien de uitslagen van deze hypothesen zou dus, o ironie, geconcludeerd mogen worden dat TO - toch een voorbeeld van de 'elaborated code' - een kenmerk vertoont van de 'restricted code'. Nader is nog onderzocht of TO ook een geringer aantal verschillende adjectieven en bijwoorden bevat. Op basis van de frequentielijsten D_1 , codes geteld naar tokens, en D_2 , codes geteld naar types (zie pagina 76), kon gemakkelijk berekend worden hoeveel verschillende adjectieven en bijwoorden per corpus voor-

komen. Hieruit bleek echter dat er aanleiding is te veronderstellen dat TO meer verschillende adjectieven en bijwoorden bevat dan D, O of W (en StG meer dan R). Verder onderzoek is hier gewenst; zie paragraaf 7.4.

7.2.4 Traditioneel

De hypothesen onder dit hoofdje worden voor het merendeel bevestigd. Alleen hypothese 41 over pronomina in verouderde naamvalsvormen springt eruit: Geen significant verschil voor D en O, terwijl voor W juist een significant verschil in niet-verwachte richting valt te constateren. Deze laatste uitslag valt te verklaren door de opvallend hoge frequentie van *der* in W. Verder is al op pagina 144 opgemerkt dat er samenhang bestaat tussen verbogen voorzetsels (hypothese 42) en substantieven in verouderde naamvalsvormen (hypothese 40), bijvoorbeeld *ten nadele van*, *ten huize van* enz.

Bij hypothese 42 kwam ook nog een verschil aan het licht bij vergelijking met het totaal aantal woorden (42) en met alleen de voorzetsels (42¹); in het laatste geval is er geen verschil tussen R en StG. Dat in de vergelijking met het totaal R meer verbogen voorzetsels bevat, komt overeen met de uitslag van hypothese 6 over voorzetsels.

En tot slot, de archaïsmen genoemd bij hypothese 43 zijn nogal willekeurig gekozen; Het enige criterium was hier, dat ze genoemd werden in publikaties over de taal van 'Den Haag'. Maar wanneer de hypothese met deze archaïsmen wordt bevestigd, dan ligt het in de rede te veronderstellen dat eenzelfde resultaat wordt bereikt bij een willekeurige andere reeks archaïsmen, bijvoorbeeld de voorzetsels *behoudens*, *betreffende* en *krachtens*, of de woorden *zomede* en *zulks*. Uiteraard moet zo'n reeks onafhankelijk van de frequentielijsten worden opgesteld; anders is er van toetsing geen sprake.

7.3 Conclusies

Na de discussie over de uitkomsten van hypothesen of aantallen samengangende hypothesen volgen in deze paragraaf de conclusies op basis van de belangrijkste verschillen tussen de onderzochte categorieën teksten.

Wanneer we de intuïtieve oordelen in hoofdstuk 3 vergelijken met de resultaten van dit kwantitatief-stilistisch onderzoek, dan mag geconcludeerd worden dat een aantal oordelen in strijd zijn met de eigenschappen van de in deze studie onderzochte steekproef uit de taal van

'Den Haag'. In tegenstelling tot wat gesuggereerd wordt in oordelen over teksten van de overheid, bevat de correspondentie tussen regering en Staten-Generaal in vergelijking met D, O of W:

1. minder 'men'
2. minder 'u', wanneer 'u' wordt vergeleken met categorieën woorden waartoe het behoort
3. minder persoonsvormen
4. minder relativa
5. minder nevenschikkende voegwoorden
6. minder voegwoorden van vergelijking
7. minder ontkennende woorden
8. minder bijvoeglijke naamwoorden
9. minder bijwoorden.

Enkele hypotheses over taalverschijnselen die niet met name genoemd worden in de oordelen, maar daaruit zijn afgeleid, worden niet bevestigd. Het betreft hier de hypotheses over:

1. substantivisch gebruikte adjectieven
2. substantivisch gebruikte deelwoorden
3. de afstand tussen delen van voornaamwoordelijke bijwoorden
4. de afstand tussen delen van samengestelde werkwoorden
5. de afstand binnen voorzetselgroepen.

Wel worden in de volgende gevallen de subjectief-impressionistische oordelen bevestigd. De correspondentie tussen regering en Staten-Generaal onderscheidt zich van de verwante norm door:

1. langere zinnen en langere woorden
2. meer lijdende vormen
3. meer (en langere) klemconstructies en meer ketens van voorzetsels en lidwoorden
4. meer substantieven en substantivisch gebruikte infinitieven, met daarbij meer substantieven op '-ing/-ingen' of in verouderde naamvalsvormen
5. meer voorzetsels
6. meer archaïsmen en verbogen voorzetsels
7. minder verwijzingen naar personen, met daarbij minder verwijzingen naar een eerste persoon.

Wanneer we binnen TO de teksten van de regering (R) en de teksten van de Staten-Generaal (StG) vergelijken, zien we dat er in de meeste gevallen geen significant verschil is aan te wijzen. Bij de acht hypotheses waar dit wel het geval is, valt op dat StG dichter staat bij D, O en W. Wanneer de hypothese luidt dat een bepaald taalverschijnsel in TO

meer voorkomt en wanneer er dan ook verschil is tussen R en StG, dan blijkt in drie gevallen het taalverschijnsel minder voor te komen in StG: voorzetsels (6), 'ondergetekende(n)' (12) en verbogen voorzetsels (42). Hypothese 19, meer persoonsvormen, is hierop slechts schijnbaar een uitzondering; StG bevat wel meer persoonsvormen dan R, maar TO heeft er minder dan D, O of W.

Hetzelfde geldt voor R in drie gevallen bij hypothesen over een taalverschijnsel dat in TO minder voorkomt. Wanneer er verschil is tussen R en StG komt het verschijnsel meer voor in StG: personalia (9), 'u' (14) en vraagzinnen (15).

De enige uitzondering op deze tendens vormt het feit dat StG meer onder-schikkende voegwoorden bevat dan R (23), terwijl TO ook meer van dit type voegwoorden bevat dan D. Maar deze uitzondering valt te verklaren wanneer we bedenken dat StG voor een groot deel uit kamervragen bestaat van één zin met meer dan één bijzin in de volgende structuur: Is de minister van mening dat ... en dat, omdat .. enz.

Het feit dat StG in de hier genoemde gevallen tussen R en D, O of W in staat, valt wellicht toe te schrijven aan de ontstaansgeschiedenis van de teksten uit de correspondentie tussen regering en Staten-Generaal. De verslagen van kamercommissies en kamervragen in StG worden als het ware sneller geschreven: er komen minder auteurs aan te pas dan bij nota's, rapporten en antwoorden op kamervragen in R, die door talrijke ambtenaren op departementen worden voorbereid. Overigens, het verschil tussen R en StG is niet groot: slechts bij 8 van de 43 hypothesen blijkt er verschil.

Het laatste onderdeel van deze paragraaf geldt de verschillen tussen D, O en W. Op basis van intuïtie zou men kunnen zeggen dat W dichterbij TO staat dan O, en O weer dichterbij D. Wanneer dit waar is, moeten - zie de gegevens op pagina 138 en 139 - bij de toetsingen die voor D, O en W een al dan niet significant verschil in verwachte richting geven, de meeste vraagtekens staan onder W en de minste onder D. Bij W zijn dit er 7; bij O 2 maar bij D ook 2. Wanneer er een al dan niet significant verschil in niet-verwachte richting optreedt, moet het omgekeerde gelden: W moet dan het laagste aantal dubbele vraagtekens hebben. Voor W is dit 1, voor O 3 en voor D 5. Deze gegevens ondersteunen de intuïtie in redelijke mate. W ligt dichterbij TO dan O of D; D ligt het verst van TO.

Maar wanneer we een aantal hypothesen afzonderlijk bekijken, zien we dat deze tendens niet in alle gevallen geldt. Bij 17, minder interjecties, en 18, minder imperatieven, had het uitroepteken dan niet onder

O maar onder D moeten staan. En bij hypothese 32, meer woorden per klemconstructie, had het vraagteken moeten gelden voor W in plaats van voor O (zie weer pagina 138 en 139).

Nog verwarrender wordt het bij toetsingsresultaten waarbij verschil aan het licht komt in verwachte richting en in niet-verwachte richting: 14 minder 'u', 30 meer 'voegwoord plus voegwoord', 41 meer verbogen pronomina. Bij 41 valt het dubbele uitroepteken onder W te verklaren door de hoge frequentie van 'der' (zie 7.2.4). Bij de andere twee is geen consistent beeld te ontdekken. Ook wanneer we ertoe zouden besluiten alleen die toetsingen te laten meetellen waarbij een taalverschijnsel wordt vergeleken met de categorie of categorieën woorden waartoe het behoort, krijgen we geen 'mooier plaatje'. Wel verandert dan hypothese 14 in 14¹, maar 18 wordt dan ook 18¹, met een (niet-significant) verschil in verwachte en niet-verwachte richting. We kunnen hier alleen concluderen dat er geen duidelijke uitspraak te doen is over 'voegwoord plus voegwoord' (30), en dat ten aanzien van 'u' slechts gesteld mag worden dat dit woord in verhouding tot de categorieën woorden waartoe het behoort, in TO significant minder voorkomt dan in D, O of W (14¹).

Als laatste manier om de verschillen tussen D, O en W in kaart te brengen, is bekeken hoe vaak toetsing een verschillende uitslag opleverde voor D, O en W. Weer op basis van intuïtie zou men kunnen zeggen, dat dit tussen D en W meer het geval moet zijn dan tussen D en O, en dat O tussen D en W ligt. De uitslag is als volgt: Het aantal verschillen tussen D en W bedraagt 13, tussen O en W 14, en tussen D en O 7. Op basis van de verschillen tussen D, O en W, voor zover onderzocht in deze studie, mag geconcludeerd worden dat D en O redelijk dicht bij elkaar liggen, en dat W de meeste overeenkomst vertoont met TO.

7.4 Verder onderzoek

Binnen het beschrijvingsmodel C_3C is er nog veel meer te onderzoeken. In deze studie zijn alleen hypothesen aan bod gekomen die zijn af te leiden uit de oordelen over de taal van 'Den Haag'. Uiteraard kan deze beschrijving buiten het C_3C -kader om verfijnd worden. Een voorbeeld: in een probabilistische grammatica, geënt op de T.G.G., zou naar aanleiding van de gegevens over relativa en klemconstructies (hypothese 21 en 31) beregeld kunnen worden hoe vaak een S uit de herschrijffregel 'NP→NP, S' getransformeerd wordt tot een relatieve bijzin of tot een klemconstructie voorafgaand aan de NP. Misschien kan deze benadering nog eens gecombineerd worden met wat *Sedelow en Sedelow* (1972:281)

voor ogen stond in hun publikatie waarin ze wijzen op het belang van het bouwen van een 'stijlmodel'.

Wanneer we ons echter beperken tot het beschrijvingsmodel in deze studie en kijken naar het overzicht van codes en frequenties (Bijlage 2), dan zijn er nog talrijke verschillen aan te wijzen tussen TO enerzijds en D, O en W anderzijds. Drie voorbeelden: De codes voor het verbogen voltooid deelwoord (207, 217, 227 en 237) komen in TO veel meer voor dan in D, O of W. Toetsing geeft hier een significant verschil. Dit verschijnsel valt onder klemconstructies, maar het is natuurlijk ook apart te onderzoeken. Het voornaamwoordelijk bijwoord (540) komt in TO meer voor dan in D. Toetsing geeft hier een significant verschil met D (niet met O en W). Eigennamen in de genitief (012) komen in TO niet voor. Op basis van Bijlage 2 kunnen alle mogelijke verschillen of clusters van verschillen gemakkelijk onderzocht worden. Wel moet hierbij worden opgemerkt, dat het ontdekken van verschillen iets anders is dan het bewijzen dat er verschillen bestaan; zie *Hemelrijk* (1958) over het verschil tussen toetsing en detectie.

Verder kunnen ook een aantal in de taalstatistiek bekende maten worden bepaald: de nomina-verba ratio, de adjectiva-nomina ratio en de type-token ratio. Uiteraard moet dan wel precies aangegeven worden wat er onder nomina enz. wordt verstaan. Bij de verba vallen bijvoorbeeld de codes voor substantivisch gebruikte infinitieven af; deze codes moeten meegeteld worden bij de nomina. Overigens is het bij het bepalen van de type-token ratio niet gewenst om uit te gaan van de frequentielijsten D_1 en D_2 (zie pagina 76), woordklassen geteld naar tokens en types. Want in D_1 geldt elke verschillende woordvorm (hoofd- of kleine letter, met en zonder afbrekingstekens) als apart type. De woordvormen, zonder deze onderscheidingen, zijn inmiddels wel beschikbaar via de retrogradelijsten (zie pagina 101).

Via de retrogradelijsten kan ook meer gedaan worden aan onderzoek naar woordvoorraad. Oordelen over de abstractiegraad van een tekst kunnen wellicht onderzocht worden door bepaalde suffixen te onderzoeken (zoals in deze studie al is gebeurd bij woorden op '-ing/-ingen'). De grote moeilijkheid blijft natuurlijk de definitie van abstracte woorden. Wanneer bijvoorbeeld in de definitie van abstracte substantieven als vormkenmerk is opgenomen het voorkomen van een van oorsprong vreemd suffix (-ie, -ica, -teit, -ade enz.) dan is zelfs 'een portie bami' abstract. Ook het oordeel 'vaagheid' zou nader onderzocht kunnen worden door naar de woordvoorraad te kijken. Men zou dan onafhankelijk van de steekproeven een lijst van woorden moeten opstellen die

als vaag (weer een definitieprobleem!) gedefinieerd worden, bijvoorbeeld samenstellingen op -aspect, -proces enz., modale bijwoorden, een deel van de indefinita uit C_3C , een reeks voorzetseluitdrukkingen als 'ten aanzien van', 'in het kader van' enz. (De voorzetseluitdrukkingen kunnen daarna via *Oog op Tekst* gemakkelijk worden opgespoord.)

Tot besluit van deze studie nog één voorbeeld van mogelijk onderzoek naar de woordvoorraad, namelijk de aardige, maar moeilijk te beantwoorden vraag: In hoeverre verschilt TO qua inhoud van D, O of W? Uiteraard moet er dan een criterium bepaald worden voor verschil in inhoud. (Een probleem voor de onderzoekers die zich met 'content-analysis' bezighouden, zie o.a. *Bereïson* (1971.) Enig inzicht op dit punt biedt een vergelijking van de tien meest frequente substantieven van elke corpus. Nuttiger zijn waarschijnlijk lijsten per corpus van substantieven of andere woordsoorten die in geen ander corpus voorkomen, maar het gaat hier alleen om een voorbeeld voor ander mogelijk onderzoek.

Op de pagina hiernaast staat de 'top-tien' van de substantieven, zoals die te vinden is in de C-lijsten, woordvormen gerangschikt naar dalende frequentie. N.B. Het gaat hier om woordvormen, dus de woorden met een kleine letter kunnen ook nog voorkomen met een hoofdletter, en omgekeerd (zie over dit punt mijn opmerkingen over de A-lijsten op pagina 76 en 77). Alleen de gelijke vormen 'jaar 000' (een jaar) en 'jaar 001' (twee jaar) zijn hier samengevoegd. Uit het overzicht blijkt dat alleen het woord 'aantal' uit TO voorkomt in D, O en W. R en StG hebben vijf woorden gemeenschappelijk. Hetzelfde geldt voor D en O, en O en W, terwijl D en W in vier woorden overeenkomen. Uiteraard is met dit voorbeeld niets bewezen over verschil of overeenkomst in inhoud. Wel is het aardig om te constateren dat in D, O en W 'mensen' heel frequent is, met in D nog 'heer' en 'minister', tegenover in TO 'leden' en 'lid'. Bovendien lijkt D iets homogener dan O en W door de vijf algemene tijdsaanduidingen. Verder lijken ook TO, R en StG homogener dan D, O of W; de frequenties van TO moeten met 2,5 vermenigvuldigd worden omdat TO kleiner is dan D, O of W. En tot slot, maar dit is intuïtie, lijkt mij de inhoud van TO vrij goed gekarakteriseerd door 'artikel' en 'vraag' enerzijds en 'regeling', 'overleg' en 'flo' (gulden) anderzijds.

Met deze impressies over woordvoorraad verlaat ik echter het terrein van deze studie. Doel van dit onderzoek was: Nagaan of subjectief-impressionistische oordelen over het taalgebruik in teksten van de overheid kunnen worden bevestigd door kwantitatief-stilistisch onderzoek. Uit de conclusies in paragraaf 7.3 blijkt dat dit slechts ten dele het geval is, en dat ook andere verschijnselen kenmerkend zijn voor de taal van 'Den Haag'.

De tien meest frequente substantieven

D	O	W
1. jaar 280	jaar 212	aantal 124
2. heer 175	mensen 146	tijd 121
3. tijd 101	tijd 112	plaats 103
4. mensen 96	land 93	mens 92
5. plaats 87	plaats 87	leven 89
6. aantal 80	jaren 85	eeuw 86
7. dag 75	Nederland 83	mensen 86
8. week 75	vraag 79	wereld 86
9. uur 74	werk 78	wijze 75
10. minister 73	aantal 75	geval 73

TO	(R)	(StG)
1. leden 126	artikel 98	leden 66
2. artikel 121	vraag 65	Minister 39
3. vraag 86	lid 64	Regering 31
4. lid 82	aantal 60	commissie 29
5. aantal 79	leden 60	artikel 23
6. aanzien 62	aanzien 51	vraag 21
7. regeling 61	overleg 48	aantal 19
8. overleg 56	regeling 47	lid 18
9. betrekking 53	betrekking 46	deel 17
10. flo 53	kader 42	geval 16

SAMENVATTING

Naar aanleiding van de talrijke kritieken op de ambtelijke stijl of de taal van 'Den Haag' is in deze studie een poging ondernomen de volgende vraag te beantwoorden: Onderscheidt het taalgebruik in teksten van de overheid zich van het taalgebruik in andere teksten, en zo ja, waarin?

Voor het beantwoorden van deze vraag zijn in de theorie van de stilistiek weinig aanknopingspunten te vinden. Maar uit de stand van zaken in de stilistiek en de talrijke verschillende stijlopvattingen blijkt wel, dat er dringend behoefte is aan beschrijvende studies om voeding te geven aan de theorie. Daarom is, ondanks zekere bezwaren, een bepaalde stijlopvatting als uitgangspunt genomen, namelijk stijl als afwijking van een (verwante) norm. Als verwante norm voor het taalgebruik in teksten van de overheid is gekozen het taalgebruik in kranten en populair-wetenschappelijke proza. De stilistische verschijnselen zijn beschreven volgens de zogenaamde kwantitatieve methode, waarbij de kwalitatieve benadering als uitgangspunt is gebruikt.

De kwalitatieve benadering vinden we terug in de zo talrijke subjectief-impressionistische oordelen over de taal van 'Den Haag'. Uit literatuur vanaf 1945 zijn zoveel mogelijk van zulke oordelen verzameld. Bij nadere beschouwing bleken deze te kunnen worden herleid tot vier trefwoorden: onpersoonlijk, ingewikkeld, wijldlopijg en traditioneel. Maar de waarde van deze oordelen blijft beperkt, onder andere omdat de gebruikte termen veelal vaag blijven. Bovendien zijn de gegeven oordelen soms tegenstrijdig.

In het hier uitgevoerde onderzoek naar aanleiding van de genoemde subjectief-impressionistische oordelen is gebruik gemaakt van een algemeen beschrijvingsmodel, namelijk het codeersysteem 'C₃C'. Dit systeem komt erop neer, dat achter elk woord in de te onderzoeken teksten een code wordt gezet met informatie over eventuele vormkenmerken en de woordsoort waartoe het woord behoort. Hoewel in deze studie kritiek wordt geleverd op dit beschrijvingsmodel en de toepassing ervan, is toch dit codeersysteem gekozen, omdat het het enige systeem is dat al op ruime schaal is toegepast op teksten die als verwante norm kunnen dienen voor de taal van 'Den Haag'. Op basis van C₃C zijn namelijk al frequentielijsten geproduceerd van grote aantallen fragmenten uit onder andere: dagbladen,

opiniebladen en populair-wetenschappelijke boeken (per categorie 120.000 woorden).

Als steekproef uit teksten van de overheid zijn fragmenten geselecteerd (in totaal 48.000 woorden) uit de correspondentie tussen regering en Staten-Generaal in het parlementaire jaar 1975-1976. Het betreft hier alleen de zogenaamde kamerstukken en kamervragen met de antwoorden daarop. Juridisch en politiek taalgebruik zijn in dit onderzoek naar ambtelijke stijl buiten beschouwing gelaten, omdat deze typen taalgebruik vermoedelijk andere kenmerken vertonen. De steekproef is gecodeerd en daarna opgenomen in het computersysteem 'Oog op Tekst'. Hierdoor werd het mogelijk zeer gedetailleerde vragen over verschillen in taalgebruik met behulp van de computer te beantwoorden. Wel bleek voor sommige vragen aanvullend handwerk noodzakelijk.

De oordelen over de taal van 'Den Haag' zijn opnieuw geformuleerd in taalkundige termen, voor zover mogelijk binnen C₃C en Oog op Tekst. Op basis hiervan zijn hypotheses opgesteld over verschillen in taalgebruik tussen teksten van de overheid enerzijds en dagbladen, opiniebladen en populair-wetenschappelijke boeken anderzijds. Ook zijn nog enkele andere hypotheses opgesteld omdat C₃C en Oog op Tekst daartoe toevallig de mogelijkheid boden. Bovendien is onderzocht of er verschillen in taalgebruik bestaan tussen teksten van de regering en teksten van de Staten-Generaal.

Toetsing van de hypotheses levert het volgende beeld op. De correspondentie tussen regering en Staten-Generaal onderscheidt zich van de verwante norm door: langere zinnen en langere woorden; meer lijdende vormen; meer en langere klemconstructies en ketens van voorzetsels en lidwoorden; meer substantieven met daarbij meer substantivisch gebruikte infinitieven en substantieven op '-ing/-ingen' of in verouderde naamvals-vormen; meer voorzetsels; meer archaïsmen en verbogen voorzetsels; minder verwijzingen naar personen met daarbij minder verwijzingen naar een eerste persoon. In al deze gevallen werden de subjectief-impressionistische oordelen over de taal van 'Den Haag' bevestigd.

Maar in tegenstelling tot wat in de oordelen wordt gesuggereerd, komen in de correspondentie tussen regering en Staten-Generaal minder persoonsvormen voor dan in de verwante norm-teksten. Hetzelfde geldt voor bijwoorden en nevenschikkende voegwoorden. Ook bij de hypotheses over 'men' en 'u' bleek toetsing juist het tegendeel op te leveren van wat op basis van de oordelen verwacht zou mogen worden, namelijk minder 'men' en meer 'u' in overheidsteksten.

Enkele resultaten die in strijd zijn met de oordelen kunnen worden verklaard door uitkomsten van andere hypotheses (bijvoorbeeld minder in plaats van meer relativa door meer klemconstructies) of door eigenschappen van de steekproef (bijvoorbeeld meer in plaats van minder vraagzinnen door het aantal kamervragen in de onderzochte teksten). Een tegenstrijdigheid in de oordelen kan worden opgelost: de zinnen zijn niet langer door meer hoofd- en bijzinnen, zoals door sommigen wordt beweerd (vergelijk de uitkomsten over persoonsvormen en voegwoorden), maar door bepalingen, zoals in andere oordelen wordt gesugereerd (vergelijk de uitkomsten over substantieven en voorzetsels).

Andere hypotheses blijken niet bevestigd te kunnen worden, bijvoorbeeld die over een grotere afstand tussen delen van samengestelde werkwoorden, meer ontkennende woorden en meer bijvoeglijke naamwoorden. Voor de laatste twee geldt dat ze juist meer voorkomen in opiniebladen en populair-wetenschappelijke boeken. Wel blijken sommige hypotheses bevestigd te kunnen worden voor bepaalde categorieën norm-teksten. De in deze studie onderzochte teksten vertonen bijvoorbeeld iets meer overeenkomst met populair-wetenschappelijke boeken. De correspondentie tussen regering en Staten-Generaal bevat namelijk meer werkwoorden (met uitzondering van onder andere verbogen deelwoorden en hulpwerkwoorden) dan dagbladen en opiniebladen, maar er blijkt geen verschil met populair-wetenschappelijke boeken; hetzelfde geldt voor eigennamen en de aanvoegende wijs.

Tussen de teksten van de regering en de teksten van de Staten-Generaal blijkt slechts een gering verschil te bestaan, in die zin dat laatstgenoemde teksten iets minder afwijken van de verwante norm.

Algemene conclusie: Dit onderzoek bevestigt slechts ten dele de gangbare oordelen over de taal van 'Den Haag' en brengt andere, al dan niet vermoede, kenmerken van dit type taalgebruik aan het licht.

THE LANGUAGE OF 'THE HAGUE'

A quantitative stylistic study on the basis of published criticisms of Dutch government language

This study, which has its raison d'être in frequent and insistent criticisms of Dutch official language (the language of 'The Hague'), attempts to provide an answer to the following question: Is the language of official publications in any way different from other, related varieties of usage, and, if so, in what respects?

Unfortunately, for an answer to this question few clues are provided by recent stylistic theory. Indeed, the state of affairs in stylistics makes it very clear that its theory is in urgent need of descriptive studies for its support and advancement. And this is the reason why, in spite of certain objections, one specific definition of style has been chosen as a working hypothesis for the present study, namely style as deviation from a related norm. As a related norm for official language, this study uses the varieties of Dutch used in newspapers and in popular scientific prose. The stylistic features discussed are defined in quantifiable terms; they have been selected on the basis of the here called 'qualitative' approach.

This qualitative approach to style we find exemplified in all those impressions, opinions, and criticisms of the language of 'The Hague' which frequently find their way into the columns of newspapers and journals. For the purpose of this study only such criticisms have been considered as were published after 1945. Closer investigation reveals that they can be categorized under four headings: the language of 'The Hague' is 'impersonal', 'complicated', 'diffuse', and 'traditional'. However, such notions are mostly left undefined, so that judgements containing them tend to be rather vague, if not actually contradictory.

For the quantitative examination and verification of hypotheses based on such impressions and opinions, this study makes use of an extant model for grammatical description, the C_3C encoding system. For every word in a text this system provides a number code which specifies its grammatical function and relevant characteristics of form. Although open to valid criticism, the C_3C system is used here in view of its being the only system which has actually been applied on a large scale, especially to such texts as can serve as a related norm to the language of 'The Hague'. On the basis of C_3C , frequency lists have

been published of the vocabulary of large samples taken from Dutch daily and weekly newspapers and popular scientific prose (120,000 words per category).

As a sample of government publications for the present study, random fragments (totalling 48,000 words) were taken from the correspondence between the government and the States General (the Dutch parliament) for the 1975-1976 parliamentary year. The sample contains parliamentary papers, including parliamentary questions and government replies. It was encoded and then stored in a computer for analysis by means of the 'Oog op Tekst' programme. 'Oog op Tekst' (lit. 'Eye on Text') is a programme especially designed for analysing texts. In this way it proved possible to answer very detailed questions on differences in vocabulary, syntax etc. For some questions, however, supplementary operations by hand were necessary.

The qualitative opinions about the language of 'The Hague' were reformulated in linguistic terms, such as fitted into C₃C and 'Oog op Tekst'. On the basis of these reformulated opinions, quantifiable hypotheses were framed concerning the differences in usage between government publications on one hand and newspapers (daily and weekly) and popular scientific publications on the other. Some additional hypotheses were suggested by specific properties of the encoding system and the computer programme. Finally, the possibility of distinctions between the language use of government and that of parliament was looked into.

The picture produced by the verification of all these hypotheses shows that the correspondence between government and States General can be distinguished from its related norm on the following points: on an average, sentences and words tend to be longer; there are more passive constructions; more words tend to be inserted between an article and its governing noun (multiple premodification); article-and-preposition chains tend to be more frequent; there are more nouns and infinitives used as nouns; nouns in '-ing' are more frequent; archaic features (certain words, case forms, inflected prepositions) are more frequent; there are fewer personal references, especially to the first person. All these findings confirmed earlier qualitative opinions.

However, the qualitative opinions were not confirmed by the following findings: government texts showed up fewer finite verbs than the related norm; there were also fewer adverbs and co-ordinating conjunctions. Rather surprisingly, verification of hypotheses concerning the use of 'men' (an indefinite personal pronoun similar to French 'on') and 'u' (the polite form of the Dutch equivalent of 'you') gave results

contrary to what was expected: 'men' was less frequent in government texts than is popularly supposed; 'u' was more frequent.

Some unexpected results can be accounted for by the outcome of tests on other hypotheses. Multiple premodification, for example, reduces the number of relative pronouns. Other surprises were due to characteristics of the sample: the relatively large frequency of interrogative constructions may have something to do with the proportion of parliamentary questions in the sample. The greater average length of sentences in these fragments, which was expected, is not due to a greater incidence of subclauses, but rather to a more frequent use of premodification and adjuncts. This explains why fewer finite verbs and conjunctions, and more nouns and prepositional constructions were found.

Other hypotheses that proved to be unfounded were the ones claiming a greater distance between the parts of a prepositional verb, a greater number of negative words, and more adjectives. These last two categories were found to be relatively more frequent in weekly newspapers and popular scientific publications. Some hypotheses could not be confirmed for all three types of text used as the related norm. The language of popular science for instance tends to show a greater similarity to the language of 'The Hague' than either type of newspaper. This is particularly true in the case of verbs (with the exception of inflected participles and auxiliaries), proper names and the use of the subjunctive.

There was found to be hardly any difference between the usage of the government and that of parliament. The States General tend to come only a little closer to the related norm.

Thus, the present study only partly confirms current opinions concerning the language of 'The Hague', and reveals other characteristics, both expected and unexpected, of this type of official language.

De eerste twee cijfers0.. als 1e cijfer: *substantieven*

0. als 2e cijfer: 'gewone' substantieven

1. " " " : eigennamen

2. " " " : adjectivisch gebruikte substantieven

8. " " " : interjectivisch gebruikte substantieven

9. " " " : woorden in zelfnoemfunctie

1.. als 1e cijfer: *adjectieven*

0. als 2e cijfer: 'gewone' adjectieven

2. " " " : substantivisch gebruikte adjectieven

5. " " " : adverbiaal gebruikte adjectieven

8. " " " : interjectivisch gebruikte adjectieven

2.. als 1e cijfer: *werkwoordsvormen*

deelwoorden en infinitieven

0. als 2e cijfer: van intransitief gebruikte werkwoorden

1. " " " : van transitief gebruikte werkwoorden

2. " " " : van reflexief gebruikte werkwoorden

3. " " " : van hulp- en koppelwerkwoorden

persoonsvormen

4. " " " : van intransitief gebruikte werkwoorden

5. " " " : van transitief gebruikte werkwoorden

6. " " " : van reflexief gebruikte werkwoorden

7. " " " : van hulp- en koppelwerkwoorden

8. " " " : van interjectivisch gebruikte werkwoorden

3.. als 1e cijfer: *pronomina A*

0. als 2e cijfer: personalia

2. " " " : possessiva (zelfstandig)

3. " " " : possessiva (bijvoeglijk)

4. " " " : reflexiva (zelfstandig)

5. " " " : reflexiva (bijvoeglijk)

6. " " " : demonstrativa (zelfstandig)

7. " " " : demonstrativa (bijvoeglijk)

- 4.. als 1e cijfer: *pronomina B*
 0. als 2e cijfer: interrogativa (zelfstandig)
 1. " " " : interrogativa (bijvoeglijk)
 2. " " " : relativa (zelfstandig)
 3. " " " : relativa (bijvoeglijk)
 4. " " " : indefinita (zelfstandig)
 5. " " " : indefinita (bijvoeglijk)
 6. " " " : cardinalia (zelfstandig)
 7. " " " : cardinalia (bijvoeglijk)
 8. " " " : ordinalia (zelfstandig)
 9. " " " : ordinalia (bijvoeglijk)
- 5.. als 1e cijfer: *bijwoorden*
 0. als 2e cijfer: 'gewone' bijwoorden
 1. " " " : aanwijzende en onbepaalde bijwoorden
 2. " " " : vragende bijwoorden
 3. " " " : betrekkelijke bijwoorden
 4. " " " : voornaamwoordelijke bijwoorden
 5. " " " : vragende voornaamwoordelijke bijwoorden
 6. " " " : betrekkelijke voornaamwoordelijke bijwoorden
 8. " " " : interjectivisch gebruikte bijwoorden
- 6.. als 1e cijfer: *pre- en postposities*
 0. als 2e cijfer: voorzetsels
 1. " " " : 2e deel gesplitste voornaamwoordelijke bijwoorden
 2. " " " : niet-verbaal deel van samengestelde werkwoorden
 3. " " " : 2e deel gescheiden voorzetselgroepen
 4. " " " : achterzetsels
 5. " " " : te bij infinitief en voorzetsel met een te+infinitief-constructie als rest
 6. " " " : voorzetsels met een door onderschikkend voegwoord ingeluide bijzin als rest
- 7.. als 1e cijfer: *conjuncties*
 0. als 2e cijfer: nevenschikkende voegwoorden
 1. " " " : onderschikkende voegwoorden
 2. " " " : voegwoorden van vergelijking
 3. " " " : onderschikkende voegwoorden die afwijkende hoofdzinsvolgorde regeren
 4. " " " : inleidend deel van nevenschikkende voegwoord-groepen
- 8.. als 1e cijfer: *interjecties*
 0. als 2e cijfer: 'echte' interjecties
 1. " " " : substantivisch gebruikte onomatopoeën

- 9.. als 1e cijfer: *restgroepen*
- 9. als 2e cijfer: ('zonder betekenis')
- 0 als 3e cijfer: *anderstalig*
- 1 " " " : niet-lexicale verbindingselementen
- 8 " " " : niet tot steekproef behorende elementen
- 9 " " " : leestekens

Het derde cijfer

Werkwoordsvormen: deelwoorden en infinitieven (2.. als eerste cijfer)

- ..0 infinitief (verbaal)
- ..1 infinitief (substantivisch)
- ..2 onvoltooid deelwoord (onverbogen)
- ..3 onvoltooid deelwoord (verbogen)
- ..4 onvoltooid deelwoord (meervoud)
- ..5 onvoltooid deelwoord (bijwoordelijk)
- ..6 voltooid deelwoord (onverbogen)
- ..7 voltooid deelwoord (verbogen)
- ..8 voltooid deelwoord (meervoud)
- ..9 voltooid deelwoord (bijwoordelijk)

Werkwoordsvormen: persoonsvormen (2.. als eerste cijfer)

- ..1 1e persoon singularis presens
- ..2 2e persoon singularis presens
- ..3 3e persoon singularis presens
- ..4 pluralis presens
- ..5 singularis preteritum
- ..6 pluralis preteritum
- ..7 imperatief zonder pronomen personale
- ..8 imperatief met pronomen personale
- ..9 overige vormen

Andere woordsoorten (eerste cijfer niet 2 en niet 9)

- ..0 basisvorm
- ..1 meervoudsvorm
- ..2 genitief
- ..3 overige verbogen vormen
- ..4 comparatief (onverbogen)
- ..5 comparatief (genitief)
- ..6 comparatief (andere verbogen vormen)
- ..7 superlatief (onverbogen)
- ..8 superlatief (genitief)
- ..9 superlatief (andere verbogen vormen)

In onderstaande tabel staat per corpus aangegeven hoe vaak een bepaalde code voorkomt. De kolommen onder D, O en W zijn overgenomen uit *Woordfrequenties*, lijst D₁ (lexicale codes, tokens). Niet vermeld zijn de codes die in deze corpora niet voorkomen: 023, 081, 268, 545 en 991. Waarschijnlijk kunnen er nog enkele codes met frequentie 1 worden geschrapt. Bijvoorbeeld, in D komt voor code 214, een transitief gebruikt onvoltooid deelwoord in het meervoud. Maar achter deze code verschuilt zich de vorm *zij maken 214 iets*; dit had moeten zijn 254. Nog een ander voorbeeld. Alleen in O komt code 235 (één maal) voor. Bij het opvragen van de zin met deze code ... *die, ouder wordend 235, zich conformeert aan het gezag* ... bleek, dat er 232 had moeten staan. Bij het vergelijken van de frequenties moet wel bedacht worden dat TO, dus R en StG samen, veel kleiner is dan D, O of W. Hieronder eerst de precieze omvang in woorden.

aantal	D	O	W	TO	(R)	(StG)
woorden	120.313	123.873	119.979	48.257	35.300	12.957

code	D	O	W	TO	(R)	(StG)
000	17717	17281	18147	8151	6013	2138
001	6571	6496	6727	3050	2256	794
002	44	31	35	4	3	1
003	84	98	134	107	80	27
010	6639	4829	2388	919	648	271
011	72	68	56	34	21	13
012	83	103	81	-	-	-
020	11	31	24	4	3	1
021	-	-	2	1	1	0
080	2	2	1	-	-	-
090	21	25	114	27	19	8
100	2072	2492	2293	823	596	227
102	13	26	12	1	1	-
103	3875	4361	4441	1554	1155	399
104	162	198	192	60	42	18
105	-	-	2	-	-	-
106	112	88	157	86	64	22
107	13	7	10	3	3	-
109	221	201	213	39	25	14

code	D	O	W	TO	(R)	(StG)
120	50	43	54	2	2	-
121	81	80	136	23	19	4
123	45	87	156	22	15	7
126	37	24	7	10	8	2
127	1	-	3	-	-	-
129	20	25	46	7	3	4
150	1737	1931	1777	404	298	106
153	11	10	34	6	5	1
154	234	263	255	79	53	26
157	54	58	50	5	4	1
159	11	11	4	5	5	-
180	3	2	-	-	-	-
200	867	936	794	248	163	85
201	123	140	205	61	47	14
202	146	103	140	41	30	11
203	424	366	463	237	185	52
204	8	2	2	8	7	1
205	38	114	123	38	26	12
206	566	498	367	269	207	62
207	51	36	34	10	9	1
208	3	3	1	2	1	1
209	31	30	15	1	1	-
210	1782	2094	1736	679	468	211
211	113	158	192	147	116	31
212	3	3	4	3	2	1
213	21	11	9	4	4	-
214	1	-	-	-	-	-
215	-	6	6	1	-	1
216	2781	2357	2435	1504	1136	368
217	322	319	383	321	256	65
218	31	40	35	12	10	2
219	7	10	38	10	9	1
220	88	102	95	36	26	10
221	1	1	4	1	1	-
222	-	-	3	-	-	-
223	-	4	4	4	4	-
224	-	1	-	-	-	-
225	-	-	3	-	-	-
226	56	57	38	9	5	4
230	1057	1055	969	729	510	219
231	5	6	9	2	2	-
232	-	1	-	-	-	-
233	3	2	6	21	10	11
234	-	-	1	-	-	-
235	-	1	-	-	-	-
236	118	146	116	29	23	6
237	1	-	-	-	-	-
241	31	65	32	1	1	-
242	19	43	9	-	-	-
243	760	1124	978	293	204	89
244	338	496	473	105	83	22
245	628	474	491	51	35	16
246	207	147	208	25	9	16
247	9	25	8	-	-	-
248	1	3	2	-	-	-
249	4	10	16	-	-	-

code	D	O	W	TO	(R)	(StG)
251	123	196	95	14	13	1
252	39	143	37	2	1	1
253	1004	1307	1255	329	239	90
254	353	594	711	126	93	33
255	955	749	612	73	26	47
256	230	187	270	59	23	36
257	31	56	45	14	11	3
258	1	2	1	-	-	-
259	5	4	7	1	1	-
261	1	9	2	1	1	-
262	1	5	1	-	-	-
263	64	67	103	15	9	6
264	40	37	53	15	10	5
265	53	36	39	-	-	-
266	19	20	20	4	1	3
267	-	4	1	-	-	-
271	164	220	96	32	31	1
272	73	153	31	16	-	16
273	3162	3427	3096	1574	1156	418
274	1165	1254	1171	571	424	147
275	1492	1318	1274	245	155	90
276	436	329	483	94	56	38
277	4	14	5	1	1	-
278	-	1	-	-	-	-
279	20	17	37	34	26	8
287	3	16	1	-	-	-
300	2049	2756	2305	253	150	103
302	-	4	2	-	-	-
303	427	618	529	66	47	19
321	1	2	1	-	-	-
323	-	2	2	-	-	-
330	1044	1249	1167	182	120	62
332	3	2	8	6	4	2
333	62	118	115	12	9	3
340	540	665	683	154	106	48
342	3	3	4	1	1	-
343	-	1	-	-	-	-
350	152	235	241	60	47	13
353	1	-	-	2	2	-
360	551	828	706	174	107	67
361	5	7	9	5	4	1
362	14	13	18	5	4	1
363	7	5	13	12	9	3
370	12901	11953	12642	6379	4668	1711
372	111	85	257	25	23	2
373	39	56	77	40	35	5
400	67	170	85	17	2	15
402	-	3	1	-	-	-
403	-	4	6	1	-	1
410	3	10	11	3	2	1
413	15	17	47	52	22	30
420	1260	1441	1381	358	272	86
422	13	11	14	2	1	1
423	10	8	37	39	33	6
430	3	3	3	1	1	-
433	2	2	3	2	1	1

code	D	O	W	TO	(R)	(StG)
440	1399	2135	1627	302	204	98
441	51	75	95	9	8	1
442	4	5	2	1	-	1
443	31	40	43	19	9	10
444	131	143	148	58	48	10
446	2	-	4	-	-	-
447	8	23	20	5	3	2
449	6	8	14	5	4	1
450	3798	4099	3915	1246	910	336
452	7	17	15	-	-	-
453	641	706	854	299	219	80
454	121	124	196	55	46	9
456	3	3	10	6	5	1
457	7	11	10	2	2	-
459	28	22	22	6	6	-
460	582	434	530	170	144	26
461	5	8	2	1	-	1
463	6	5	4	-	-	-
470	1874	1096	889	740	572	168
471	30	21	16	3	2	1
473	3	3	-	-	-	-
480	19	16	22	4	3	1
481	4	6	3	2	2	-
483	7	13	23	1	1	-
490	170	90	158	52	49	3
493	157	113	123	73	63	10
500	5503	6334	5895	1701	1190	511
503	6	4	2	-	-	-
504	8	6	6	1	-	1
507	-	1	1	-	-	-
509	-	1	2	-	-	-
510	1867	2512	2235	520	358	162
520	118	181	169	62	22	40
530	92	114	118	20	15	5
540	673	837	763	352	265	87
550	9	17	10	-	-	-
560	359	349	348	152	110	42
580	4	15	4	-	-	-
600	15600	14542	14622	7819	5879	1940
603	132	134	194	220	184	36
610	8	4	2	-	-	-
620	4	3	1	-	-	-
630	68	98	68	8	6	2
640	5	15	7	2	2	-
650	1664	1823	1588	688	485	203
660	15	11	14	3	3	-
700	3262	4057	4131	1040	775	265
710	1692	2223	2050	872	584	288
720	643	823	903	253	183	70
730	23	43	30	14	10	4
740	32	49	59	35	25	10
800	59	143	49	19	10	9
810	-	-	1	-	-	-
990	12	22	151	18	15	3

1. substantieven (+)

In TO komen de codes 000, 001, 020 en 021 meer voor dan in D, O of W.

	D	TO	
Subgroep	24299 0,2020	11206 0,2322	35505 Afwijking in verwachte richting
Restant	96014 0,7980	37051 0,7678	133065 Overschrijdingskans: 0,00000000
	120313	48257	168570

	O	TO	
Subgroep	23808 0,1922	11206 0,2322	35014 Afwijking in verwachte richting
Restant	100065 0,8078	37051 0,7678	137116 Overschrijdingskans: 0,00000000
	123873	48257	172130

	W	TO	
Subgroep	24900 0,2075	11206 0,2322	36106 Afwijking in verwachte richting
Restant	95079 0,7925	37051 0,7678	132130 Overschrijdingskans: 0,00000000
	119979	48257	168236

	R	StG	
Subgroep	8273 0,2344	2933 0,2264	11206
Restant	27027 0,7656	10024 0,7736	37051 Overschrijdingskans: 0,06694982
	35300	12957	48257

2. substantivisch gebruikte adjectieven (+)

In TO komen de codes 120, 121, 123, 126, 127 en 129 meer voor dan in D, O of W.

	D	TO	
Subgroep	234 0,0019	64 0,0013	298 Afwijking in niet-verwachte richting
Restant	120079 0,9981	48193 0,9987	168272 Overschrijdingskans: 0,00760378
	120313	48257	168570
	O	TO	
Subgroep	259 0,0021	64 0,0013	323 Afwijking in niet-verwachte richting
Restant	123614 0,9979	48193 0,9987	171807 Overschrijdingskans: 0,00123591
	123873	48257	172130
	W	TO	
Subgroep	402 0,0034	64 0,0013	466 Afwijking in niet-verwachte richting
Restant	119577 0,9966	48193 0,9987	167770 Overschrijdingskans: 0,00000000
	119979	48257	168236
	R	StG	
Subgroep	47 0,0013	17 0,0013	64
Restant	35253 0,9987	12940 0,9987	48193 Overschrijdingskans 0,92893123
	35300	12957	48257

3. substantivisch gebruikte infinitieven (+)

In TO komen de codes 201, 211, 221 en 231 meer voor dan in D, O of W.

	D	TO	
Subgroep	242 0,0020	211 0,0044	453 Afwijking in verwachte richting
Restant	120071 0,9980	48046 0,9956	168117 Overschrijdingskans: 0,00000000
	120313	48257	168570

	O	TO	
Subgroep	305 0,0025	211 0,0044	516 Afwijking in verwachte richting
Restant	123568 0,9975	48046 0,9956	171614 Overschrijdingskans: 0,00000000
	123873	48257	172130

	W	TO	
Subgroep	410 0,0034	211 0,0044	621 Afwijking in verwachte richting
Restant	119569 0,9966	48046 0,9956	167615 Overschrijdingskans: 0,00400897
	119979	48257	168236

	R	StG	
Subgroep	166 0,0047	45 0,0035	211
Restant	35134 0,9953	12912 0,9965	48046 Overschrijdingskans: 0,08249901
	35300	12957	48257

4. substantivisch gebruikte deelwoorden (+)

In TO komen de codes 204, 214, 224, 234, 208, 218 en de substantivisch gebruikte deelwoorden in verbogen vorm, meer voor dan in D, O of W.

	D	TO	
Subgroep	59 0,0005	36 0,0007	95 Afwijking in verwachte richting
Restant	120254 0,9995	48221 0,9993	168475 Overschrijdingskans: 0,05938003
	120313	48257	168570

	O	TO	
Subgroep	75 0,0006	36 0,0007	111 Afwijking in verwachte richting
Restant	123798 0,9994	48221 0,9993	172019 Overschrijdingskans: 0,35442255
	123873	48257	172130

	W	TO	
Subgroep	68 0,0006	36 0,0007	104 Afwijking in verwachte richting
Restant	119911 0,9994	48221 0,9993	168132 Overschrijdingskans: 0,21894242
	119979	48257	168236

	R	StG	
Subgroep	31 0,0009	5 0,0004	36
Restant	35269 0,9991	12952 0,9996	48221 Overschrijdingskans: 0,11704600
	35300	12957	48257

5. werkwoorden (-)

In TO komen de codes 2.. met tweede cijfer 0, 1 of 2 (en daarna alleen ..0, ..2 of ..6) en met tweede cijfer 4, 5 of 6 (en daarna geen ..7, ..8 of ..9) minder voor dan in D, 0 of W.

	D	TO	
Subgroep	12329 0,1025	4660 0,0966	16989 Afwijking in verwachte richting
Restant	107984 0,8975	43597 0,9034	151581 Overschrijdingskans: 0,00027985
	120313	48257	168570

	0	TO	
Subgroep	13051 0,1054	4660 0,0966	17711 Afwijking in verwachte richting
Restant	110822 0,8946	43597 0,9034	154419 Overschrijdingskans: 0,00000007
	123873	48257	172130

	W	TO	
Subgroep	12086 0,1007	4660 0,0966	16746 Afwijking in verwachte richting
Restant	107893 0,8993	43597 0,9034	151490 Overschrijdingskans: 0,01006220
	119979	48257	168236

	R	StG	
Subgroep	3318 0,0940	1342 0,1036	4660
Restant	31982 0,9060	11615 0,8964	43597 Overschrijdingskans: 0,00168982
	35300	12957	48257

6. voorzetsels (+)

In TO komen de codes 600 en 603 meer voor dan in D, O of W.

	D	TO	
Subgroep	15732 0,1308	8039 0,1666	23771 Afwijking in verwachte richting
Restant	104581 0,8692	40218 0,8334	144799 Overschrijdingskans: 0,00000000
	120313	48257	168570

	O	TO	
Subgroep	14676 0,1185	8039 0,1666	22715 Afwijking in verwachte richting
Restant	109197 0,8815	40218 0,8334	149415 Overschrijdingskans: 0,00000000
	123873	48257	172130

	W	TO	
Subgroep	14816 0,1235	8039 0,1666	22855 Afwijking in verwachte richting
Restant	105163 0,8765	40218 0,8334	145381 Overschrijdingskans: 0,00000000
	119979	48257	168236

	R	StG	
Subgroep	6063 0,1718	1976 0,1525	8039
Restant	29237 0,8282	10981 0,8475	40218 Overschrijdingskans: 0,00000053
	35300	12957	48257

7. substantieven op '-ing' of '-ingen' (+)

In T0 komen woorden op '-ing' of '-ingen' meer voor dan in D, O of W.

	D	T0		
Subgroep	2552 0,0212	2386 0,0494	4938	Afwijking in verwachte richting
Restant	117761 0,9788	45871 0,9506	163632	Overschrijdingskans: 0,00000000
	120313	48257	168570	

	O	T0		
Subgroep	2332 0,0188	2386 0,0494	4718	Afwijking in verwachte richting
Restant	121541 0,9812	45871 0,9506	167412	Overschrijdingskans: 0,00000000
	123873	48257	172130	

	W	T0		
Subgroep	2678 0,0223	2386 0,0494	5064	Afwijking in verwachte richting
Restant	117301 0,9777	45871 0,9506	163172	Overschrijdingskans: 0,00000000
	119979	48257	168236	

	R	StG		
Subgroep	1809 0,0512	577 0,0445	2386	
Restant	33491 0,9488	12380 0,9555	45871	Overschrijdingskans: 0,00277501
	35300	12957	48257	

8. lijdende vorm (+)

In TO komt de lijdende vorm meer voor dan in D, O of W.

	D	TO	
Subgroep	392 0,0110	1128 0,0234	1520 Afwijking in verwachte richting
Restant	35100 0,9890	47129 0,9766	82229 Overschrijdingskans: 0,00000000
	35492	48257	83749

	O	TO	
Subgroep	384 0,0108	1128 0,0234	1512 Afwijking in verwachte richting
Restant	35168 0,9892	47129 0,9766	82297 Overschrijdingskans: 0,00000000
	35552	48257	83809

	W	TO	
Subgroep	492 0,0121	1128 0,0234	1620 Afwijking in verwachte richting
Restant	40301 0,9879	47129 0,9766	87430 Overschrijdingskans: 0,00000000
	40793	48257	89050

	R	StG	
Subgroep	869 0,0246	259 0,0200	1128
Restant	34431 0,9754	12698 0,9800	47129 Overschrijdingskans: 0,00319535
	35300	12957	48257

9. personalia (-)

In TO komen personalia die verwijzen naar personen, een gedeelte van de codes 300, 302 en 303, minder voor dan in D, O of W.

	D	TO	
Subgroep	2251 0,0187	234 0,0048	2485 Afwijking in verwachte richting
Restant	118062 0,9813	48023 0,9952	166085 Overschrijdingskans: 0,00000000
	120313	48257	168570

	O	TO	
Subgroep	3124 0,0252	234 0,0048	3358 Afwijking in verwachte richting
Restant	120749 0,9748	48023 0,9952	168772 Overschrijdingskans: 0,00000000
	123873	48257	172130

	W	TO	
Subgroep	2434 0,0203	234 0,0048	2668 Afwijking in verwachte richting
Restant	117545 0,9797	48023 0,9952	165568 Overschrijdingskans: 0,00000000
	119979	48257	168236

	R	StG	
Subgroep	137 0,0039	97 0,0075	234
Restant	35163 0,9961	12860 0,9925	48023 Overschrijdingskans: 0,00000064
	35300	12957	48257

10. possessiva (-)

In T0 komen possessiva die verwijzen naar personen, een gedeelte van de codes 321, 323, 330, 332 en 333, minder voor dan in D, O of W.

	D	T0	
Subgroep	931 0,0077	116 0,0024	1047 Afwijking in verwachte richting
Restant	119382 0,9923	48141 0,9976	167523 Overschrijdingskans: 0,00000000
	120313	48257	168570

	O	T0	
Subgroep	1212 0,0098	116 0,0024	1328 Afwijking in verwachte richting
Restant	122661 0,9902	48141 0,9976	170802 Overschrijdingskans: 0,00000000
	123873	48257	172130

	W	T0	
Subgroep	1019 0,0085	116 0,0024	1135 Afwijking in verwachte richting
Restant	118960 0,9915	48141 0,9976	167101 Overschrijdingskans: 0,00000000
	119979	48257	168236

	R	StG	
Subgroep	77 0,0022	39 0,0030	116
Restant	35223 0,9978	12918 0,9970	48141 Overschrijdingskans: 0,12294055
	35300	12957	48257

11. verwijzingen naar een 'eerste persoon' (-)

In TO komen verwijzingen naar een 'eerste persoon' minder voor dan in D, O of W.

	D	TO	
Subgroep	1051 0,0087	124 0,0026	1175 Afwijking in verwachte richting
Restant	119262 0,9913	48133 0,9974	167395 Overschrijdingskans: 0,00000000
	120313	48257	168570

	O	TO	
Subgroep	1736 0,0140	124 0,0026	1860 Afwijking in verwachte richting
Restant	122137 0,9860	48133 0,9974	170270 Overschrijdingskans: 0,00000000
	123873	48257	172130

	W	TO	
Subgroep	1382 0,0115	124 0,0026	1506 Afwijking in verwachte richting
Restant	118597 0,9885	48133 0,9974	166730 Overschrijdingskans: 0,00000000
	119979	48257	168236

	R	StG	
Subgroep	95 0,0027	29 0,0022	124
Restant	35205 0,9973	12928 0,9978	48133 Overschrijdingskans: 0,44143254
	35300	12957	48257

12. 'ondergetekende(n)' (+)

In T0 komt het woord 'ondergetekende(n)' meer voor dan in D, O of W.

	D	T0	
Subgroep	0 0,0000	69 0,0014	69 Afwijking in verwachte richting
Restant	120313 1,0000	48188 0,9986	168501 Overschrijdingskans: 0,00000000
	120313	48257	168570

	O	T0	
Subgroep	0 0,0000	69 0,0014	69 Afwijking in verwachte richting
Restant	123873 1,0000	48188 0,9986	172061 Overschrijdingskans: 0,00000000
	123873	48257	172130

	W	T0	
Subgroep	0 0,0000	69 0,0014	69 Afwijking in verwachte richting
Restant	119979 1,0000	48188 0,9986	168167 Overschrijdingskans: 0,00000000
	119979	48257	168236

	R	StG	
Subgroep	69 0,0020	0 0,0000	69
Restant	35231 0,9980	12957 1,0000	48188 Overschrijdingskans: 0,00000096
	35300	12957	48257

13. 'men' (+)

In TO komt het woord 'men' meer voor dan in D, O of W.

	D	TO	
Subgroep	182 0,0015	36 0,0007	218 Afwijking in niet-verwachte richting
Restant	120131 0,9985	48221 0,9993	168352 Overschrijdingskans: 0,00010259
	120313	48257	168570

	O	TO	
Subgroep	317 0,0026	36 0,0007	353 Afwijking in niet-verwachte richting
Restant	123556 0,9974	48221 0,9993	171777 Overschrijdingskans: 0,00000000
	123873	48257	172130

	W	TO	
Subgroep	448 0,0037	36 0,0007	484 Afwijking in niet-verwachte richting
Restant	119531 0,9963	48221 0,9993	167752 Overschrijdingskans: 0,00000000
	119979	48257	168236

	R	StG	
Subgroep	22 0,0006	14 0,0011	36
Restant	35278 0,9994	12943 0,9989	48221 Overschrijdingskans: 0,14919114
	35300	12957	48257

14. 'u' (-)

In T0 komt het woord 'u' minder voor dan in D, O of W.

	D	T0	
Subgroep	73 0,0006	33 0,0007	106
Restant	120240 0,9994	48224 0,9993	168464
	120313	48257	168570

Afwijking in niet-verwachte richting

Overschrijdingskans: 0,64320287

	O	T0	
Subgroep	132 0,0011	33 0,0007	165
Restant	123741 0,9989	48224 0,9993	171965
	123873	48257	172130

Afwijking in verwachte richting

Overschrijdingskans: 0,02694735

	W	T0	
Subgroep	64 0,0005	33 0,0007	97
Restant	119915 0,9995	48224 0,9993	168139
	119979	48257	168236

Afwijking in niet-verwachte richting

Overschrijdingskans: 0,29366356

	R	StG	
Subgroep	11 0,0003	22 0,0017	33
Restant	35289 0,9997	12935 0,9983	48224
	35300	12957	48257

Overschrijdingskans: 0,00000068

15. vraagzinnen (-)

In TO komen vraagzinnen minder voor dan in D, O of W.

	D	TO	
Subgroep	64 0,0097	147 0,0752	211 Afwijking in niet-verwachte richting
Restant	6567 0,9903	1808 0,9248	8375 Overschrijdingskans: 0,00000000
	6631	1955	8586

	O	TO	
Subgroep	209 0,0307	147 0,0752	356 Afwijking in niet-verwachte richting
Restant	6603 0,9693	1808 0,9248	8411 Overschrijdingskans: 0,00000000
	6812	1955	8767

	W	TO	
Subgroep	117 0,0203	147 0,0752	264 Afwijking in niet-verwachte richting
Restant	5635 0,9797	1808 0,9248	7443 Overschrijdingskans: 0,00000000
	5752	1955	7707

	R	StG	
Subgroep	1 0,0007	146 0,2709	147
Restant	1415 0,9993	393 0,7291	1808 Overschrijdingskans: 0,00000000
	1416	539	1955

16. eigennamen (-)

In T0 komen de codes 010, 011 en 012 minder voor dan in D, O of W.

	D	T0	
Subgroep	6794 0,0565	953 0,0197	7747 Afwijking in verwachte richting
Restant	113519 0,9435	47304 0,9803	160823 Overschrijdingskans: 0,00000000
	120313	48257	168570

	O	T0	
Subgroep	5000 0,0404	953 0,0197	5953 Afwijking in verwachte richting
Restant	118873 0,9596	47304 0,9803	166177 Overschrijdingskans: 0,00000000
	123873	48257	172130

	W	T0	
Subgroep	2525 0,0210	953 0,0197	3478 Afwijking in verwachte richting
Restant	117454 0,9790	47304 0,9803	164758 Overschrijdingskans: 0,09453320
	119979	48257	168236

	R	StG	
Subgroep	669 0,0190	284 0,0219	953
Restant	34631 0,9810	12673 0,9781	47304 Overschrijdingskans: 0,04144745
	35300	12957	48257

17. interjecties (-)

In TO komen de codes 080, 180, 287, 580, 800 en 810 minder voor dan in D, O of W.

	D	TO	
Subgroep	71 0,0006	19 0,0004	90 Afwijking in verwachte richting
Restant	120242 0,9994	48238 0,9996	168480 Overschrijdingskans: 0,14394330
	120313	48257	168570

	O	TO	
Subgroep	178 0,0014	19 0,0004	197 Afwijking in verwachte richting
Restant	123695 0,9986	48238 0,9996	171933 Overschrijdingskans: 0,00000001
	123873	48257	172130

	W	TO	
Subgroep	56 0,0005	19 0,0004	75 Afwijking in verwachte richting
Restant	119923 0,9995	48238 0,9996	168161 Overschrijdingskans: 0,60720989
	119979	48257	168236

	R	StG	
Subgroep	10 0,0003	9 0,0007	19
Restant	35290 0,9997	12948 0,9993	48238 Overschrijdingskans: 0,07847359
	35300	12957	48257

18. imperatieven (-)

In TO komen de codes 247, 248, 257, 258, 267, 277 en 278 minder voor dan in D, O of W.

	D	TO	
Subgroep	46 0,0004	15 0,0003	61 Afwijking in verwachte richting
Restant	120267 0,9996	48242 0,9997	168509 Overschrijdingskans: 0,57819065
	120313	48257	168570
	O	TO	
Subgroep	105 0,0008	15 0,0003	120 Afwijking in verwachte richting
Restant	123768 0,9992	48242 0,9997	172010 Overschrijdingskans: 0,00022565
	123873	48257	172130
	W	TO	
Subgroep	62 0,0005	15 0,0003	77 Afwijking in verwachte richting
Restant	119917 0,9995	48242 0,9997	168159 Overschrijdingskans: 0,09691116
	119979	48257	168236
	R	StG	
Subgroep	12 0,0003	3 0,0002	15
Restant	35288 0,9997	12954 0,9998	48242 Overschrijdingskans: 0,75856044
	35300	12957	48257

19. persoonsvormen (+)

In T0 komen de codes 24., 25., 26. en 27. meer voor dan in D, O of W.

	D	T0	
Subgroep	11432 0,0950	3695 0,0766	15127
Restant	108881 0,9050	44562 0,9234	153443
	120313	48257	168570

Afwijking in niet-verwachte
richting

Overschrijdingskans: 0,00000000

	O	T0	
Subgroep	12536 0,1012	3695 0,0766	16231
Restant	111337 0,8988	44562 0,9234	155899
	123873	48257	172130

Afwijking in niet-verwachte
richting

Overschrijdingskans: 0,00000000

	W	T0	
Subgroep	11662 0,0972	3695 0,0766	15357
Restant	108317 0,9028	44562 0,9234	152879
	119979	48257	168236

Afwijking in niet-verwachte
richting

Overschrijdingskans: 0,00000000

	R	StG	
Subgroep	2609 0,0739	1086 0,0838	3695
Restant	32691 0,9261	11871 0,9162	44562
	35300	12957	48257

Overschrijdingskans: 0,00030898

20. zinslengte (+)

In TO zijn de zinnen over het algemeen langer dan in D, O of W.

	D	TO	O	TO
1.	19,13	24,76	19,17	24,76
2.	0,12	0,31	0,15	0,31
3.	0,30	0,80	0,38	0,80
4.	18,83 - 19,43	23,96 - 25,56	18,79 - 19,55	23,96 - 25,56

	W	TO	(R)	(StG)
1.	21,44	24,76	25,17	23,68
2.	0,17	0,31	0,34	0,53
3.	0,45	0,80	0,88	1,38
4.	20,99 - 21,89	23,96 - 25,56	24,29 - 26,05	22,30 - 25,00

1. gemiddelde
2. standaarddeviatie gedeeld door de wortel van het aantal waarnemingen
3. 2,58 maal '2'
4. het betrouwbaarheidsinterval voor het gemiddelde in de populatie

De standaarddeviatie is een maat voor de spreiding van de waarnemingen rond het gemiddelde. Hoe kleiner de standaarddeviatie, des te minder variatie is er in de waarnemingen. Gesteld dat de standaarddeviatie 0 zou zijn, dan is elke zin even lang.

Deze maat is nodig om significantieverschillen tussen gemiddelden op te sporen. (Zie ook de hypothesen 27, 28, 29, 32 en 36.)

21. relativa (+)

In TO komen de codes 420, 422, 423, 430 en 433 meer voor dan in D, O of W.

	D	TO	
Subgroep	1288 0,0107	402 0,0083	1690 Afwijking in niet-verwachte richting
Restant	119025 0,9893	47855 0,9917	166880 Overschrijdingskans: 0,00001096
	120313	48257	168570

	O	TO	
Subgroep	1465 0,0118	402 0,0083	1867 Afwijking in niet-verwachte richting
Restant	122408 0,9882	47855 0,9917	170263 Overschrijdingskans: 0,00000000
	123873	48257	172130

	W	TO	
Subgroep	1438 0,0120	402 0,0083	1840 Afwijking in niet-verwachte richting
Restant	118541 0,9880	47855 0,9917	166396 Overschrijdingskans: 0,00000000
	119979	48257	168236

	R	StG	
Subgroep	308 0,0087	94 0,0073	402
Restant	34992 0,9913	12863 0,9927	47855 Overschrijdingskans: 0,12887913
	35300	12957	48257

22. betrekkelijke (voornaamwoordelijke) bijwoorden (+)

In TO komen de codes 530 en 560 meer voor dan in D, O of W.

	D	TO	
Subgroep	451 0,0037	172 0,0036	623 Afwijking in niet-verwachte richting
Restant	119862 0,9963	48085 0,9964	167947 Overschrijdingskans: 0,60356166
	120313	48257	168570

	O	TO	
Subgroep	463 0,0037	172 0,0036	635 Afwijking in niet-verwachte richting
Restant	123410 0,9963	48085 0,9964	171495 Overschrijdingskans: 0,62491040
	123873	48257	172130

	W	TO	
Subgroep	466 0,0039	172 0,0036	638 Afwijking in niet-verwachte richting
Restant	119513 0,9961	48085 0,9964	167598 Overschrijdingskans: 0,35691560
	119979	48257	168236

	R	StG	
Subgroep	125 0,0035	47 0,0036	172
Restant	35175 0,9965	12910 0,9964	48085 Overschrijdingskans: 0,95628771
	35300	12957	48257

23. onderschikkende voegwoorden (+)

In T0 komen de codes 710 en 730 meer voor dan in D, O of W.

	D	T0		
Subgroep	1715 0,0143	886 0,0184	2601	Afwijking in verwachte richting
Restant	118598 0,9857	47371 0,9816	165969	Overschrijdingskans: 0,00000000
	120313	48257	168570	

	O	T0		
Subgroep	2266 0,0183	886 0,0184	3152	Afwijking in verwachte richting
Restant	121607 0,9817	47371 0,9816	168978	Overschrijdingskans: 0,94160384
	123873	48257	172130	

	W	T0		
Subgroep	2080 0,0173	886 0,0184	2966	Afwijking in verwachte richting
Restant	117899 0,9827	47371 0,9816	165270	Overschrijdingskans: 0,15487488
	119979	48257	168236	

	R	StG		
Subgroep	594 0,0168	292 0,0225	886	
Restant	34706 0,9832	12665 0,9775	47371	Overschrijdingskans: 0,00004101
	35300	12957	48257	

24. te 650 (+)

In TO komt het woord 'te' met code 650 meer voor dan in D, O of W.

	D	TO	
Subgroep	1245 0,0103	550 0,0114	1795 Afwijking in verwachte richting
Restant	119068 0,9897	47707 0,9886	166775 Overschrijdingskans: 0,06134215
	120313	48257	168570

	O	TO	
Subgroep	1386 0,0112	550 0,0114	1936 Afwijking in verwachte richting
Restant	122487 0,9888	47707 0,9886	170194 Overschrijdingskans: 0,73168456
	123873	48257	172130

	W	TO	
Subgroep	1214 0,0101	550 0,0114	1764 Afwijking in verwachte richting
Restant	118765 0,9899	47707 0,9886	166472 Overschrijdingskans: 0,02129563
	119979	48257	168236

	R	StG	
Subgroep	392 0,0111	158 0,0122	550
Restant	34908 0,9889	12799 0,9878	47707 Overschrijdingskans: 0,34173464
	35300	12957	48257

25. nevenschikkende voegwoorden (+)

In TO komen de codes 700 en 740 meer voor dan in D, O of W.

	D	TO	
Subgroep	3294 0,0274	1075 0,0223	4369 Afwijking in niet-verwachte richting
Restant	117019 0,9726	47182 0,9777	164201 Overschrijdingskans: 0,00000000
	120313	48257	168570

	O	TO	
Subgroep	4106 0,0331	1075 0,0223	5181 Afwijking in niet-verwachte richting
Restant	119767 0,9669	47182 0,9777	166949 Overschrijdingskans: 0,00000000
	123873	48257	172130

	W	TO	
Subgroep	4190 0,0349	1075 0,0223	5265 Afwijking in niet-verwachte richting
Restant	115789 0,9651	47182 0,9777	162971 Overschrijdingskans: 0,00000000
	119979	48257	168236

	R	StG	
Subgroep	800 0,0227	275 0,0212	1075
Restant	34500 0,9773	12682 0,9788	47182 Overschrijdingskans: 0,36052695
	35300	12957	48257

26. voegwoorden van vergelijking (+)

In TO komt code 720 meer voor dan in D, O of W.

	D	TO	
Subgroep	643 0,0053	253 0,0052	896
Restant	119670 0,9947	48004 0,9948	167674
	120313	48257	168570

Afwijking in niet-verwachte richting

Overschrijdingskans: 0,82404590

	O	TO	
Subgroep	823 0,0066	253 0,0052	1076
Restant	123050 0,9934	48004 0,9948	171054
	123873	48257	172130

Afwijking in niet-verwachte richting

Overschrijdingskans: 0,00104237

	W	TO	
Subgroep	903 0,0075	253 0,0052	1156
Restant	119076 0,9925	48004 0,9948	167080
	119979	48257	168236

Afwijking in niet-verwachte richting

Overschrijdingskans: 0,00000035

	R	StG	
Subgroep	183 0,0052	70 0,0054	253
Restant	35117 0,9948	12887 0,9946	48004
	35300	12957	48257

Overschrijdingskans: 0,82334962

27. afstand tussen delen van voornaamwoordelijke bijwoorden (+)

In TO is de afstand tussen code 610 en het bijbehorende woord over het algemeen groter dan in D, O of W.

	D	TO	O	TO
1.	2,16	1,43	2,26	1,43
2.	0,12	0,20	0,10	0,20
3.	0,32	0,51	0,25	0,51
4.	1,94 - 2,58	0,92 - 1,94	2,02 - 2,51	0,92 - 1,94

	W	TO	(R)	(StG)
1.	2,10	1,43	1,16	2,10
2.	0,11	0,20	0,23	0,31
3.	0,29	0,51	0,59	0,81
4.	1,81 - 2,39	0,92 - 1,94	0,57 - 1,75	1,29 - 2,91

1. gemiddelde
2. standaarddeviatie gedeeld door de wortel van het aantal waarnemingen
3. 2,58 maal '2'
4. het betrouwbaarheidsinterval voor het gemiddelde in de populatie

28. afstand tussen delen van samengestelde werkwoorden (+)

In TO is de afstand tussen code 620 en het bijbehorende woord over het algemeen groter dan in D, O of W.

	D	TO	O	TO
1.	3,00	2,50	2,58	2,50
2.	0,14	0,23	0,11	0,23
3.	0,36	0,59	0,28	0,59
4.	2,64 - 3,36	1,91 - 3,09	2,30 - 2,86	1,91 - 3,09

	W	TO	(R)	(StG)
1.	2,76	2,50	2,45	2,60
2.	0,13	0,23	0,28	0,42
3.	0,34	0,59	0,72	1,08
4.	2,42 - 3,10	1,91 - 3,09	1,73 - 3,17	1,52 - 3,68

1. gemiddelde
2. standaarddeviatie gedeeld door de wortel van het aantal waarnemingen
3. 2,58 maal '2'
4. het betrouwbaarheidsinterval voor het gemiddelde in de populatie

29. afstand binnen voorzetselgroepen (+)

In TO is de afstand tussen code 630 en het bijbehorende woord over het algemeen groter dan in D, O of W.

	D	TO	O	TO
1.	1,47	1,13	1,85	1,13
2.	0,16	0,13	0,19	0,13
3.	0,42	0,32	0,50	0,32
4.	1,05 - 1,89	0,80 - 1,45	1,35 - 2,35	0,80 - 1,45

	W	TO	(R)	(StG)
1.	1,65	1,13	1	1,5
2.	0,16	0,13	0	0,5
3.	0,42	0,32	0	1,29
4.	1,23 - 2,07	0,80 - 1,45	1	0,21 - 2,79

1. gemiddelde
2. standaarddeviatie gedeeld door de wortel van het aantal waarnemingen
3. 2,58 maal '2'
4. het betrouwbaarheidsinterval voor het gemiddelde in de populatie

30. voegwoord plus voegwoord (+)

In T0 komt de combinatie 7...-7... meer voor dan in D, O of W.

	D	T0	
Subgroep	56 0,0005	21 0,0004	77 Afwijking in niet-verwachte richting
Restant	120257 0,9995	48236 0,9996	168493 Overschrijdingskans: 0,89108582
	120313	48257	168570

	O	T0	
Subgroep	72 0,0006	21 0,0004	93 Afwijking in niet-verwachte richting
Restant	123801 0,9994	48236 0,9996	172037 Overschrijdingskans: 0,29099410
	123873	48257	172130

	W	T0	
Subgroep	50 0,0004	21 0,0004	71 Afwijking in verwachte richting
Restant	119929 0,9996	48236 0,9996	168165 Overschrijdingskans: 0,97188647
	119979	48257	168236

	R	StG	
Subgroep	10 0,0003	11 0,0008	21
Restant	35290 0,9997	12946 0,9992	48236 Overschrijdingskans: 0,01665275
	35300	12957	48257

31. klemconstructies met deelwoord of infinitief (+)

In TO is het aantal klemconstructies met deelwoord of infinitief groter dan in D, O of W.

	D	TO	
Subgroep	717 0,0060	540 0,0112	1257 Afwijking in verwachte richting
Restant	119596 0,9940	47717 0,9888	167313 Overschrijdingskans: 0,00000000
	120313	48257	168570

	O	TO	
Subgroep	617 0,0050	540 0,0112	1157 Afwijking in verwachte richting
Restant	123256 0,9950	47717 0,9888	170973 Overschrijdingskans: 0,00000000
	123873	48257	172130

	W	TO	
Subgroep	729 0,0061	540 0,0112	1269 Afwijking in verwachte richting
Restant	119250 0,9939	47717 0,9888	166967 Overschrijdingskans: 0,00000000
	119979	48257	168236

	R	StG	
Subgroep	421 0,0119	119 0,0092	540
Restant	34879 0,9881	12838 0,9908	47717 Overschrijdingskans: 0,01280792
	35300	12957	48257

32. aantal woorden per klemconstructie (+)

In TO is het aantal woorden per klemconstructie met deelwoord of infinitief over het algemeen groter dan in D, O of W.

	D	TO	O	TO
1.	1,11	1,55	1,27	1,55
2.	0,06	0,09	0,08	0,09
3.	0,15	0,23	0,21	0,23
4.	0,96 - 1,26	1,32 - 1,78	1,06 - 1,48	1,32 - 1,78

	W	TO	(R)	(StG)
1.	1,04	1,55	1,61	1,34
2.	0,06	0,09	0,11	0,17
3.	0,16	0,23	0,28	0,44
4.	0,88 - 1,20	1,32 - 1,78	1,33 - 1,89	0,90 - 1,78

1. gemiddelde
2. standaarddeviatie gedeeld door de wortel van het aantal waarnemingen
3. 2,58 maal '2'
4. het betrouwbaarheidsinterval voor het gemiddelde in de populatie

33. ketens van voorzetsels en lidwoorden (+)

In TO komen de ketens 'voorzetsel-lidwoord-voorzetsel-lidwoord' en 'lidwoord-voorzetsel-lidwoord-voorzetsel' meer voor dan in D, O of W.

	D	TO	
Subgroep	25 0,0002	52 0,0011	77 Afwijking in verwachte richting
Restant	120288 0,9998	48205 0,9989	168493 Overschrijdingskans: 0,00000000
	120313	48257	168570

	O	TO	
Subgroep	21 0,0002	52 0,0011	73 Afwijking in verwachte richting
Restant	123852 0,9998	48205 0,9989	172057 Overschrijdingskans: 0,00000000
	123873	48257	172130

	W	TO	
Subgroep	26 0,0002	52 0,0011	78 Afwijking in verwachte richting
Restant	119953 0,9998	48205 0,9989	168158 Overschrijdingskans: 0,00000000
	119979	48257	168236

	R	StG	
Subgroep	44 0,0012	8 0,0006	52
Restant	35256 0,9988	12949 0,9994	48205 Overschrijdingskans: 0,08725977
	35300	12957	48257

34. ontkennende woorden (+)

In TO komen ontkennende woorden meer voor dan in D, O of W.

	D	TO	
Subgroep	1229 0,0102	452 0,0094	1681
Restant	119084 0,9898	47805 0,9906	166889
	120313	48257	168570

Afwijking in niet-verwachte
richting
Overschrijdingskans: 0,11929995

	O	TO	
Subgroep	1720 0,0139	452 0,0094	2172
Restant	122153 0,9861	47805 0,9906	169958
	123873	48257	172130

Afwijking in niet-verwachte
richting
Overschrijdingskans: 0,00000000

	W	TO	
Subgroep	1375 0,0115	452 0,0094	1827
Restant	118604 0,9885	47805 0,9906	166409
	119979	48257	168236

Afwijking in niet-verwachte
richting
Overschrijdingskans: 0,00019783

	R	StG	
Subgroep	314 0,0089	138 0,0107	452
Restant	34986 0,9911	12819 0,9893	47805
	35300	12957	48257

Overschrijdingskans: 0,08527600

35. woorden met 'niet-' 'non-' of 'on-' (+)

In TO komen woorden met 'niet-', 'non-' of 'on-' meer voor dan in D, O of W.

	D	TO	
Subgroep	304 0,0025	98 0,0020	402 Afwijking in niet-verwachte richting
Restant	120009 0,9975	48159 0,9980	168168 Overschrijdingskans: 0,06698184
	120313	48257	168570

	O	TO	
Subgroep	355 0,0029	98 0,0020	453 Afwijking in niet-verwachte richting
Restant	123518 0,9971	48159 0,9980	171677 Overschrijdingskans: 0,00283554
	123873	48257	172130

	W	TO	
Subgroep	360 0,0030	98 0,0020	458 Afwijking in niet-verwachte richting
Restant	119619 0,9970	48159 0,9980	167778 Overschrijdingskans: 0,00067176
	119979	48257	168236

	R	StG	
Subgroep	72 0,0020	26 0,0020	98
Restant	35228 0,9980	12931 0,9980	48159 Overschrijdingskans: 0,96596543
	35300	12957	48257

36. woordlengte (+)

In TO zijn de woorden over het algemeen langer dan in D, O of W.

	D	TO	O	TO
1.	5,39	5,57	5,28	5,57
2.	0,01	0,02	0,01	0,02
3.	0,03	0,05	0,03	0,05
4.	5,37 - 5,42	5,52 - 5,62	5,25 - 5,30	5,52 - 5,62

	W	TO	(R)	(StG)
1.	5,32	5,57	5,56	5,59
2.	0,01	0,02	0,02	0,03
3.	0,03	0,05	0,05	0,09
4.	5,30 - 5,35	5,52 - 5,62	5,51 - 5,61	5,50 - 5,68

1. gemiddelde
2. standaarddeviatie gedeeld door de wortel van het aantal waarnemingen
3. 2,58 maal '2'
4. het betrouwbaarheidsinterval voor het gemiddelde in de populatie

37. adjectieven (+)

In TO komen de codes 100, 102, 103, 104, 105, 106, 107 en 109 meer voor dan in D, O of W.

	D	TO	
Subgroep	6468 0,0538	2566 0,0532	9034 Afwijking in niet-verwachte richting
Restant	113845 0,9462	45691 0,9468	159536 Overschrijdingskans: 0,63760293
	120313	48257	168570

	O	TO	
Subgroep	7373 0,0595	2566 0,0532	9939 Afwijking in niet-verwachte richting
Restant	116500 0,9405	45691 0,9468	162191 Overschrijdingskans: 0,00000042
	123873	48257	172130

	W	TO	
Subgroep	7320 0,0610	2566 0,0532	9886 Afwijking in niet-verwachte richting
Restant	112659 0,9390	45691 0,9468	158350 Overschrijdingskans: 0,00000000
	119979	48257	168236

	R	StG	
Subgroep	1886 0,0534	680 0,0525	2566
Restant	33414 0,9466	12277 0,9475	45691 Overschrijdingskans: 0,69818265
	35300	12957	48257

38. bijwoorden (+)

In TO komen de codes 150, 153, 154, 157, 159, 500, 503, 504, 507 en 509 meer voor dan in D, O of W.

	D	TO	
Subgroep	7564 0,0629	2201 0,0456	9765 Afwijking in niet-verwachte richting
Restant	112749 0,9371	46056 0,9544	158805 Overschrijdingskans: 0,00000000
	120313	48257	168570

	O	TO	
Subgroep	8619 0,0696	2201 0,0456	10820 Afwijking in niet-verwachte richting
Restant	115254 0,9304	46056 0,9544	161310 Overschrijdingskans: 0,00000000
	123873	48257	172130

	W	TO	
Subgroep	8026 0,0669	2201 0,0456	10227 Afwijking in niet-verwachte richting
Restant	111953 0,9331	46056 0,9544	158009 Overschrijdingskans: 0,00000000
	119979	48257	168236

	R	StG	
Subgroep	1555 0,0441	646 0,0499	2201
Restant	33745 0,9559	12311 0,9501	46056 Overschrijdingskans: 0,00725935
	35300	12957	48257

39. aanvoegende wijs (+)

In TO komen de codes 249, 259 en 279 meer voor dan in D, O of W.

	D	TO	
Subgroep	29 0,0002	35 0,0007	64 Afwijking in verwachte richting
Restant	120284 0,9998	48222 0,9993	168506 Overschrijdingskans: 0,00000765
	120313	48257	168570

	O	TO	
Subgroep	31 0,0003	35 0,0007	66 Afwijking in verwachte richting
Restant	123842 0,9997	48222 0,9993	172064 Overschrijdingskans: 0,00001162
	123873	48257	172130

	W	TO	
Subgroep	60 0,0005	35 0,0007	95 Afwijking in verwachte richting
Restant	119919 0,9995	48222 0,9993	168141 Overschrijdingskans: 0,09995043
	119979	48257	168236

	R	StG	
Subgroep	27 0,0008	8 0,0006	35
Restant	35273 0,9992	12949 0,9994	48222 Overschrijdingskans: 0,73202415
	35300	12957	48257

40. substantieven in verouderde naamvalsvormen (+)

In TO komen code 003 en een deel van code 002 meer voor dan in D, O of W.

	D	TO	
Subgroep	94 0,0008	111 0,0023	205 Afwijking in verwachte richting
Restant	120219 0,9992	48146 0,9977	168365 Overschrijdingskans: 0,00000000
	120313	48257	168570

	O	TO	
Subgroep	108 0,0009	111 0,0023	219 Afwijking in verwachte richting
Restant	123765 0,9991	48146 0,9977	171911 Overschrijdingskans: 0,00000000
	123873	48257	172130

	W	TO	
Subgroep	156 0,0013	111 0,0023	267 Afwijking in verwachte richting
Restant	119823 0,9987	48146 0,9977	167969 Overschrijdingskans: 0,00000438
	119979	48257	168236

	R	StG	
Subgroep	83 0,0024	28 0,0022	111
Restant	35217 0,9976	12929 0,9978	48146 Overschrijdingskans: 0,77986851
	35300	12957	48257

41. pronomina in verouderde naamvalsvormen (+)

In TO komen pronomina in verouderde naamvalsvormen meer voor dan in D, O of W.

	D	TO	
Subgroep	116 0,0010	53 0,0011	169 Afwijking in verwachte richting
Restant	120197 0,9990	48204 0,9989	168401 Overschrijdingskans: 0,48301361
	120313	48257	168570

	O	TO	
Subgroep	109 0,0009	53 0,0011	162 Afwijking in verwachte richting
Restant	123764 0,9991	48204 0,9989	171968 Overschrijdingskans: 0,21515649
	123873	48257	172130

	W	TO	
Subgroep	304 0,0025	53 0,0011	357 Afwijking in niet-verwachte richting
Restant	119675 0,9975	48204 0,9989	167879 Overschrijdingskans: 0,00000001
	119979	48257	168236

	R	StG	
Subgroep	46 0,0013	7 0,0005	53
Restant	35254 0,9987	12950 0,9995	48204 Overschrijdingskans: 0,03686816
	35300	12957	48257

42. verbogen voorzetsels (+)

In T0 komt code 603 meer voor dan in D, O of W.

	D	T0		
Subgroep	132 0,0011	220 0,0046	352	Afwijking in verwachte richting
Restant	120181 0,9989	48037 0,9954	168218	Overschrijdingskans: 0,00000000
	120313	48257	168570	

	O	T0		
Subgroep	134 0,0011	220 0,0046	354	Afwijking in verwachte richting
Restant	123739 0,9989	48037 0,9954	171776	Overschrijdingskans: 0,00000000
	123873	48257	172130	

	W	T0		
Subgroep	194 0,0016	220 0,0046	414	Afwijking in verwachte richting
Restant	119785 0,9984	48037 0,9954	167822	Overschrijdingskans: 0,00000000
	119979	48257	168236	

	R	StG		
Subgroep	184 0,0052	36 0,0028	220	
Restant	35116 0,9948	12921 0,9972	48037	Overschrijdingskans: 0,00057875
	35300	12957	48257	

43. enkele archaïsmen (+)

In TO komen de woorden *alsmede, alsook, betreffende 103, hetgeen, hetwelk, middels, onderhavig(e), uiteraard, waar 710, wat ... betreft* en *welk(e)* 42./43. meer voor dan in D, O of W.

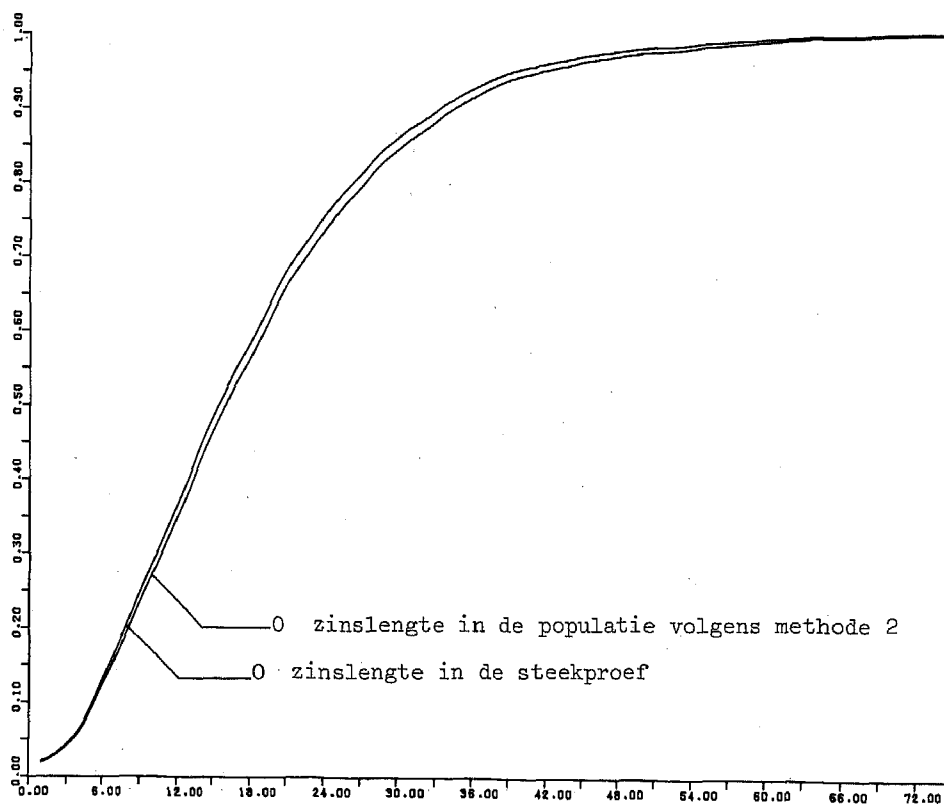
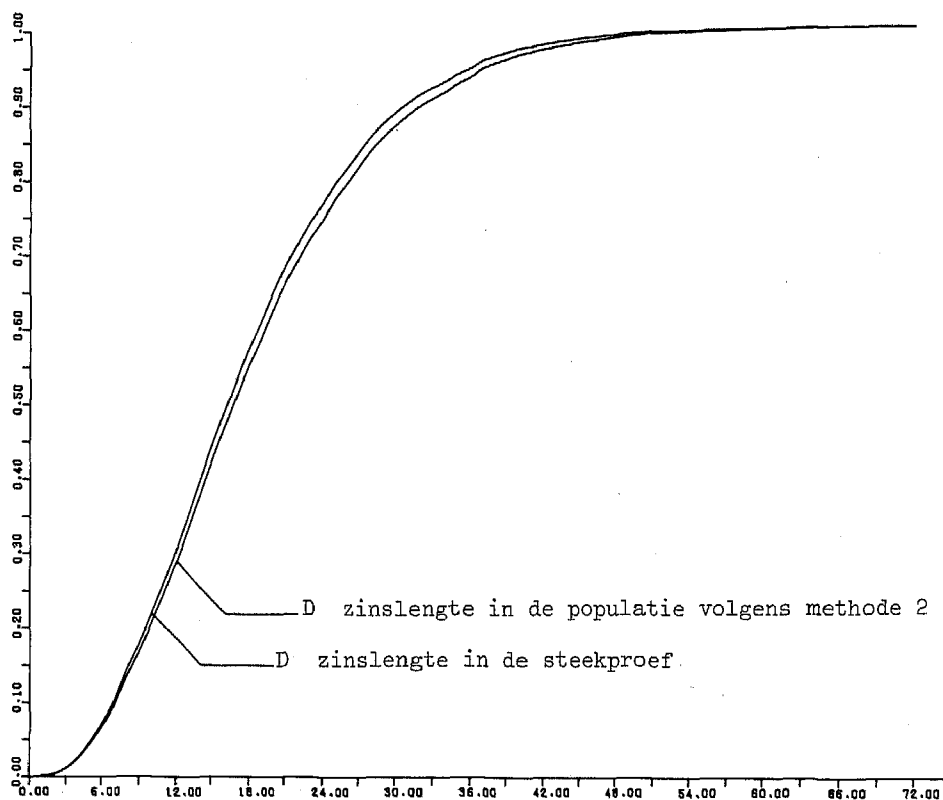
	D	TO	
Subgroep	56 0,0005	139 0,0029	195 Afwijking in verwachte richting
Restant	120257 0,9995	48118 0,9971	168375 Overschrijdingskans: 0,00000000
	120313	48257	168570

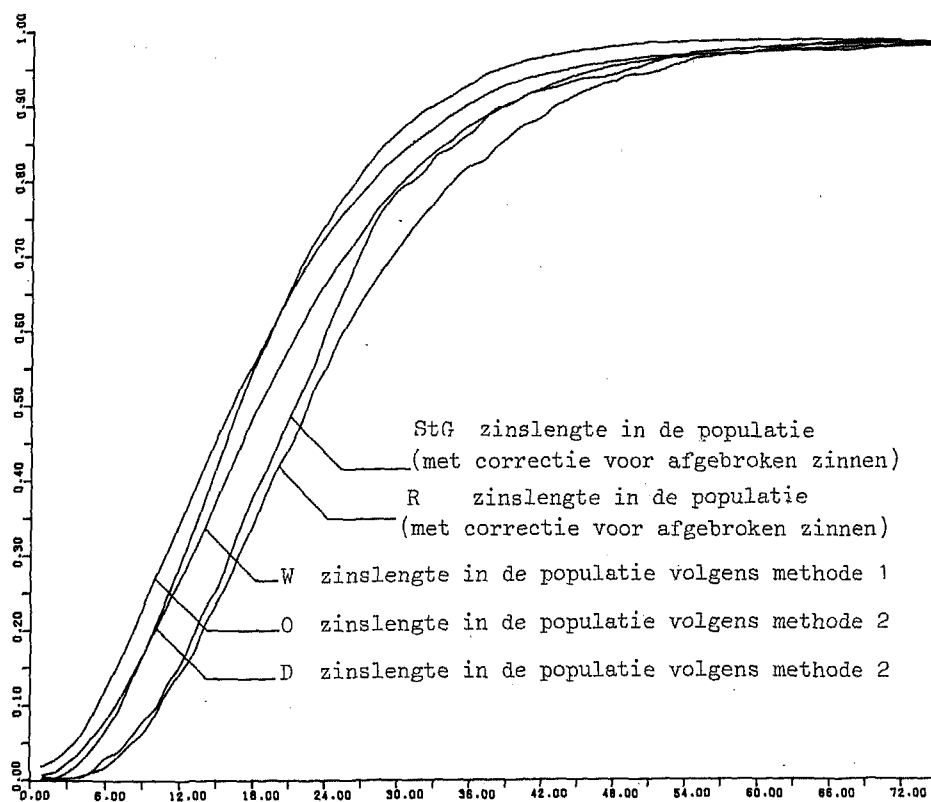
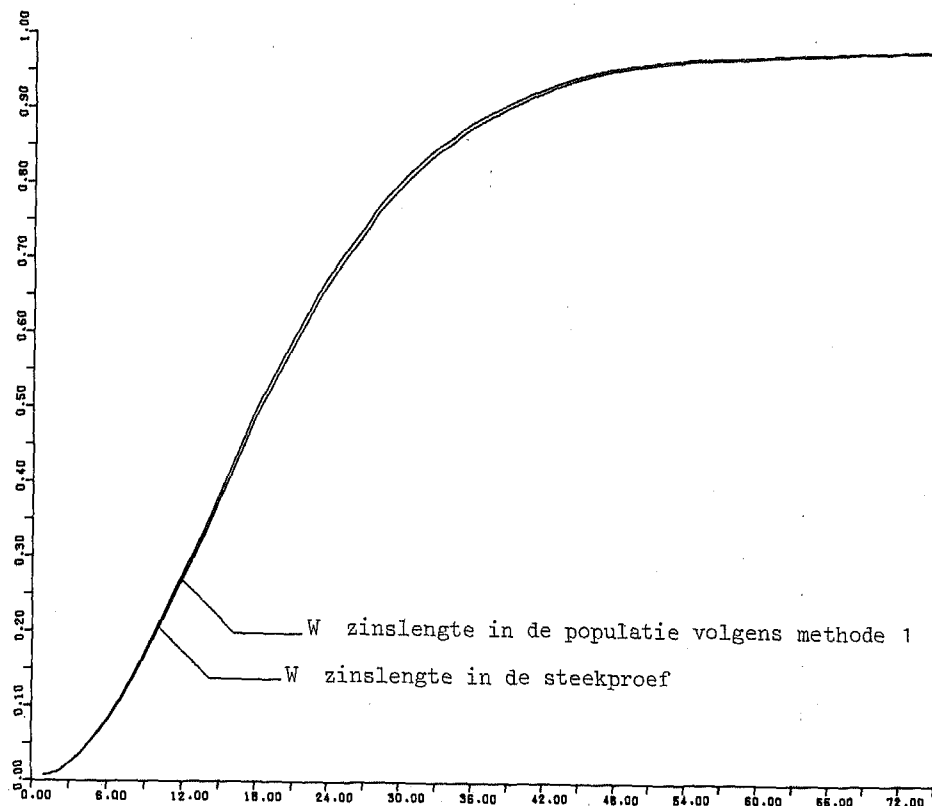
	O	TO	
Subgroep	75 0,0006	139 0,0029	214 Afwijking in verwachte richting
Restant	123798 0,9994	48118 0,9971	171916 Overschrijdingskans: 0,00000000
	123873	48257	172130

	W	TO	
Subgroep	133 0,0011	139 0,0029	272 Afwijking in verwachte richting
Restant	119846 0,9989	48118 0,9971	167964 Overschrijdingskans: 0,00000000
	119979	48257	168236

	R	StG	
Subgroep	110 0,0031	29 0,0022	139
Restant	35190 0,9969	12928 0,9978	48118 Overschrijdingskans: 0,13384909
	35300	12957	48257

GRAFIEKEN ZINSLENGTE





LITERATUUR

(Voor aangehaalde publikaties staan paginaverwijzingen. Een + geeft aan dat de publikatie is gebruikt bij het verzamelen van oordelen over de taal van 'Den Haag'.)

- 37 + *Aanwijzingen inzake indeling en stijl van ambtelijke stukken*, vastgesteld bij besluit van de minister-president van 4 september 1970, 's-Gravenhage.
Abraham, W., 'Stil, Pragmatik und Abweichungsgrammatik', A. von Stechow (Hrsg.), *Beiträge zur generativen Grammatik*, Braunschweig, 1971, pag. 1-13.
Akhmanova, O., *Linguostylistics, Theory and Method*, The Hague, 1976.
Al, B.P.F., 'Norm, afwijking en interpretatie', *Forum der Letteren* (20), 1979, pag. 420-427.
Antosch, F., 'The Diagnosis of Literary Style with the Verb-Adjective Ratio', L. Doležal and R.W. Bailey (eds.), *Statistics and Style*, New York, 1969, pag. 57-65.
- 8, 156 Appel, R., Hubers, G. en Meijer, G., *Sociolinguïstiek*, Utrecht/Antwerpen, 1979 (tweede druk).
- 1 Assen, C.J. van, *De taal der Grondwet en eenige aantekeningen*, eerste stuk, Leiden, 1844.
Bailey, R.W., 'Statistics and Style: a Historical Survey', L. Doležal and R.W. Bailey (eds.), *Statistics and Style*, New York, 1969, pag. 217-236.
Bailey, R.W., 'Toward the Integrity of Stylistics: Symbiosis vs. Parasitism', B.B. Kachru and H.F.W. Stahlke (eds.), *Current Trends in Stylistics*, Edmonton, 1972, pag. 97-101.
Bailey, R.W., Doležal, L., *An Annotated Bibliography of Statistical Stylistics*, Ann Arbor, 1968.
- 34 + Bakker, J.J.M., 'Hoe schrijf ik Expertinees', *Onze Taal* (33), 1964, pag. 15-16.
- 26 Beauvais, R., *L'Hexagonale tel qu'on parle*, Paris, 1970.
Bennett, P.E., 'The Statistical Measurement of a Stylistic Trait in Julius Ceasar and As You Like It', L. Doležal and R.W. Bailey (eds.), *Statistics and Style*, New York, 1969, pag. 29-41.
Berckel, J.A.Th. van, e.a., *Formal Properties of Newspaper Dutch*, Amsterdam, 1965.
- 162 Berelson, B., *Content Analysis in Communication Research*, New York, 1971 (eerste druk 1952).
Berg, B. van de, 'W. Martin. De inhoud van krant en roman; een frequentie-onderzoek' (boekbeoordeling), *De Nieuwe Taalgids* (62), 1969, pag. 139-140.
Boot, M., 'Een model voor automatische woordklaskodering', *Handelingen van het XXXIe Vlaams Filologencongres*, Brussel, 1977, pag. 225-234.
- 30, 35 + Braam, A. van, *Ambtenaren en bureaucratie in Nederland*, 's-Gravenhage, 40, 43 1957.

- Brandt Corstius, H., 'Taalstatistiek', *Statistica Neerlandica* (4), 1964, pag. 353-367.
- Brandt Corstius, H., *Weer een nieuw vak? Taken en grenzen van de computer-taalkunde*, Rotterdam, 1974.
- Brandt Corstius, H., 'Lange woorden', *Spektator* (7), 1977, pag. 48-52.
- 70, 100 Brandt Corstius, H., *Computer-taalkunde*, Muiderberg, 1978.
- Buch, K.R., 'A Note on Sentence-Length as Random Variable', L. Doležal and R.W. Bailey (eds.), *Statistics and Style*, New York, 1969, pag. 76-79.
- Carroll, J.B., 'Vectors of Prose Style', T.A. Sebeok (ed.), *Style in Language*, Cambridge Mass., 1975, pag. 283-292 (zesde druk).
- 12 Carstensen, B., 'Stil und Norm', *Zeitschrift für Dialektologie und Linguistik* (37), 1970, pag. 257-279.
- Chan, R.K., *Illustration of a Methodology for Stylistic Investigation with Special Reference to Formality in English*, University of Reading, 1970.
- Codeervoorschrift, zie: Voorschriften lexicale codering in het systeem C₃C.
- Court, J.F.H.A. de la, *De meest voorkomende woorden en woordcombinaties in het Nederlandsch*, Batavia, 1937.
- 7, 11 Crystal, D., 'Objective and Subjective in Stylistic Analysis', B.B. Kachru and H.F.W. Stahlke (eds.), *Current Trends in Stylistics*, Edmonton, 1972, pag. 103-114.
- 7, 23 Crystal, D., Davy, D., *Investigating English Style*, London, 1969.
- 33 + Van Dale, *Groot Woordenboek der Nederlandse Taal*, 's-Gravenhage, 1976 (tiende druk).
- David, J., Martin, R. (eds.), *Statistique et linguistique*, Colloque organisé par le centre d'analyse syntaxique de l'université de Metz, Paris, 1974.
- Doležal, L., 'A Framework for the Statistical Analysis of Style', L. Doležal and R.W. Bailey (eds.), *Statistics and Style*, New York, 1969, pag. 10-25.
- Doležal, L., Kraus, J., 'Prague School Stylistics', B.B. Kachru and H.F.W. Stahlke (eds.), *Current Trends in Stylistics*, Edmonton, 1972, pag. 37-48.
- 26, 39 + Drop, W., Vries, J.H.L. de, *Taalbeheersing, handboek voor taalhantering*, Groningen 1974.
- 39 + Drop, W., Vries, J.H.L. de, *Taalbeheersing, handboek voor taalhantering*, Groningen, 1977 (tweede, gewijzigde druk).
- Drijkoningen, F.F.J., 'Stilistiek en het onderzoek der periodestijlen', *Forum der Letteren* (4), 1963, pag. 207-214.
- Ducrot, P.R., 'Quantitative Stylistics: an Essay in Methodology', *Computers and the Humanities* (IV), 1970, pag. 187-191.
- Dyen, I., *Linguistic Subgrouping and Lexicostatistics*, The Hague, 1975.
- Eggers, H., 'Nachdenkliches über Aktiv und Passiv', *Muttersprache* (89), 1979, pag. 47-52.
- Ellis, J.M., 'Linguistics, Literature and the Concept of Style', *Word* (26), 1970, pag. 65-78.
- 25 + Elsevier, *Onze taal verloedert*, 10 juni 1972.

- 7, 9, 12, 26 Enkvist, N.E., 'On Defining Style, an Essay in Applied Linguistics', J. Spencer (ed.), *Linguistics and Style*, London, 1964, pag. 1-56.
- 10, 13 Enkvist, N.E., *Linguistic Stylistics*, The Hague, 1973.
- Es, G.A. van, Caspel, P.P.J. van, *Publikaties van het archief voor de Nederlandse syntaxis*, Reeks I, nr. 23, Passieve Constructie I; nr. 24, Passieve Constructie II, Groningen, 1973.
- Faik, S., 'Neutralisation de phénomènes parasites dans les calculs de fréquence', *Etudes de linguistique appliquée* (6), 1972, pag. 16-27.
- Fischer, W.L., 'Topologische Stilcharakteristiken von Texten', *Grundlagenstudien aus Kybernetik und Geisteswissenschaft* (10), 1969, pag. 111-119.
- Fischer, W.L., 'Beispiele für topologische Stilcharakteristiken von Texten', *Grundlagenstudien aus Kybernetik und Geisteswissenschaft* (11), 1970, pag. 1-11.
- Fleischer, W., 'Zur funktionalstilistischen Differenzierung der deutschen Schriftsprache', *Sprachpflege* (18), 1969, pag. 225-230.
- Fowler, R. (ed.), *Essays on Style and Language. Linguistic and Critical Approaches to Literary Style*, London, 1970.
- Fowler, R., 'Style and the Concept of Deep Structure', *Journal of Literary Semantics* (1), 1972, pag. 5-24.
- 12 Freeman, D.C., 'Linguistic Approaches to Literature', D.C. Freeman (ed.), *Linguistics and Literary Style*, New York, 1970, pag. 3-18.
- Fucks, W., *Mathematische Analyse von Sprachelementen, Sprachstil und Sprachen*, Köln, 1953.
- Fucks, W., Lauter, J., 'Mathematische Analyse des literarischen Stils', H. Kreuzer en R. Gunzenhäuser (Hrsg.), *Mathematik und Dichtung*, München, 1971, pag. 107-122 (vierde druk).
- 68 Gerritsen, M., 'Woordgetel', *Meta* (10), 1976, pag. 6-10.
- 68 Gerritsen, M., 'Een kwantitatief onderzoek naar sexeverschillen in het Amsterdams', G. Geerts en A. Hagen (red.), *Sociolinguïstische studies*, I, Groningen, 1980, pag. 154-179.
- Gibson, W., 'Styles and Statistics. A Model T Style Machine', D.C. Freeman (ed.), *Linguistics and Literary Style*, New York, 1970, pag. 143-164.
- 146 Gill, R., 'Correction of Systematic Errors in Sentence Sampling', (rapport van het Mathematisch Centrum), Amsterdam, te verschijnen.
- Ginneken, J. van, 'De statistiek en de taalwetenschap', *De Nieuwe Taalgids* (9), 1915, pag. 65-95.
- Ginneken, J. van, 'De woordfrequentie', *Onze Taaltuin* (7), 1939, pag. 289-313.
- Gläser, R., 'Die Kategorie "Funktionalstil" in soziolinguïstischer Sicht', *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* (27), 1974, pag. 487-496.
- 26 Gowers, E., *The Complete Plain Words*, Harmondsworth, 1973.
- 10, 17 Gray, B., *Style, the Problem and its Solution*, The Hague, 1969.
- Gregory, M., 'Aspects of Varieties Differentiation', *Journal of Linguistics* (3), 1967, pag. 177-198.
- Gregory, M., Carroll, S., *Language and Situation; Language Varieties and their Social Contexts*, London, 1978.
- 6 Groot, A.D. de, *Methodologie, grondslagen van onderzoek en denken in de gedragswetenschappen*, 's-Gravenhage, 1968 (vierde druk).

- 17, 8 Guiraud, P., 'La Stylistique' (*Que sais je?* nr. 646), Paris, 1954a (vijfde druk 1967).
- Guiraud, P., *Bibliographie Critique de la Statistique Linguistique*, Utrecht 1954b. (Revisée et complétée par Thomas D. Houchin, Jaan Puhvel et Calvert W. Watkins sous la direction de Joshua Whatmough.)
- Guiraud, P., *Les caractères statistiques du vocabulaire, Essai de méthodologie*, Paris, 1954c.
- Guiraud, P., *Problèmes et méthodes de la statistique linguistique*, Paris, 1960.
- Gülich, E., Raible, W. (Hrsg.), *Textsorten. Differenzierungskriterien aus linguistischer Sicht*, Frankfurt, 1972.
- Haeringen, C.B. van, '"Tangconstructies" en reacties daarop', *De Nieuwe Taalgids* (40), 1947, pag. 1-7.
- Haeringen, C.B. van, 'Het achtervoegsel -ing: mogelijkheden en beperkingen', *De Nieuwe Taalgids* (64), 1971, pag. 449-468.
- 8 Haft-van Rees, A., 'Register', *Spektator* (3), 1974, pag. 433-446.
- 36, 38 + Hamer, G.W., *Mensentaal*, Werk- en discussieboek voor taaltraining en
40, 43 -vorming bij de Provincie Overijssel, Zwolle, 1975.
- 8 Hartemink, F., Rijlaarsdam, G., *Register, een theoretisch model experimenteel getoetst*, Leiden, 1978 (doctoraalscriptie).
- Hasan, R., 'Code, Register and Social Dialect', B. Bernstein (ed.), *Class, codes and control* (vol. 2), London, 1975, pag. 253-292 (tweede druk).
- 8 Hartveldt, D., *Taal en samenleving. Over de sociale functies van taal, ideologie en taalvariatie*, Baarn, 1978.
- Hartveldt, D., 'Situatie en taalgebruik: een kwestie van interpretatie', *Levende Talen* (338/339), 1979a, pag. 30-34.
- Hartveldt, D., 'De functie van een chaos', *Levende Talen* (344), 1979b, pag. 625-632.
- 36, 38, 127 + Hauwermeiren, P. van, *Het leesbaarheidsonderzoek*, Groningen, 1975.
- Hauwermeiren, P. van, 'Enkele kwantitatieve aspecten van de Nederlandse romantaal', *Bijdragen over Letterkunde en Estetiek*, Hoger Instituut voor Vertalers en Tolken, Gent, z.j.
- Hauwermeiren, P. van, 'Leesbaar Nederlands (XI), concrete en specifieke -rende stijl', *Taalbeheersing in de Administratie* (171), 1979, pag. I, 720 - I, 722.
- Havránek, B., 'The Functional Differentiation of the Standard Language', P.L. Garvin (ed.), *A Prague School Reader on Esthetics, Literary Structure, and Style*, Washington D.C., 1964, pag. 3-16.
- Hayes, C.W., 'A Study in Prose Styles: Edward Gibbon and Ernest Hemingway', *Texas Studies in Literature and Language*, VII, 1966, pag. 371-386. (Een verkorte versie in L. Doležal and R.W. Bailey eds., *Statistics and Style*, New York, 1969, pag. 80-91; ook gepubliceerd in D.C. Freeman (ed.), *Linguistics and Literary Style*, New York, 1970, pag. 279-296.)
- Hellings, W.Gs., Merwe Scholtz, H. van der, *Kreatieve analyse van taalgebruik. Prinsipes van stilistiek op linguïstische grondslag*, Amsterdam, 1955.
- 161 Hemelrijk, J., 'Statistische proefopzetten: bewijs en detectie', *Statistica Neerlandica* (12), 1958, pag. 111-118.
- 135, 136 Hemelrijk, J., *Oriënterende cursus mathematische statistiek* (interne publikatie van het Mathematisch Centrum Amsterdam, syllabus 34), Amsterdam, 1977.

- 21 Herdan, G., *Language as Choice and Chance*, Groningen, 1956.
- 21 Herdan, G., *The Calculus of Linguistic Observations*, 's-Gravenhage, 1962.
- Herdan, G., 'Quantitative Linguistics or Generative Grammar', *Linguistics* (13), 1964, pag. 56-65.
- 21 Herdan, G., *The Advanced Theory of Language as Choice and Chance*, Berlin, 1966.
- 26, 34, 35, 36 + Hermkens, H.M., *Verzorgd Nederlands*, Den Bosch, 1974 (vijfde druk).
38, 39, 42
- 148 Hertog, C.H. den, *Nederlandse Spraakkunst*, derde stuk, De leer van de woordsoorten. Ingeleid en bewerkt door H. Hulshof, Amsterdam, 1973.
- 68 Heuven, V.J.J.P. van, *Spelling en Lezen. Hoe tragisch zijn de werkwoordsvormen?*, Assen, 1978
- Horálek, K., 'Sprachfunktion und funktionelle Stilistik', *Linguistics* (14), 1965, pag. 14-22.
- 81 Jaarverslag Staatsdrukkerij- en Uitgeverijbedrijf, Den Haag, 1975.
- 84, 89 Jaarverslag Staatsdrukkerij- en Uitgeverijbedrijf, Den Haag, 1976.
- Jacobs, R.A., Rosenbaum, P.S., *Transformations, Style, and Meaning*, Waltham, Mass., 1971.
- Jakobson, R., 'Closing Statement: Linguistics and Poetics', T.A. Sebeok (ed.), *Style in Language*, Cambridge Mass., 1975, pag. 350-377 (zesde druk).
- 26, 35 + Japin, L.M., *Taal en techniek, Praktisch taalgebruik voor technici*, Culemborg, 1968 (tweede druk).
- 47 Jong, E.D. de (red.), *Spreektaal. Woordfrequenties in gesproken Nederlands*, Utrecht, 1979.
- Joos, M., 'The Five Clocks', *International Journal of American Linguistics* (28), 1962, pag. 7-62.
- 31, 40 + Kamp, A.F., 'Het ambtelijk taalgebruik', *Onze Taal* (30), 1961, pag. 8-11.
- Kerkhoff, E.L., *Het begrip stijl*, Groningen, 1946.
- Kerkhoff, E.L., *De kunst der stijlinterpretatie, een pleidooi voor functionele stilistiek*, Amsterdam, 1951.
- Kirsner, R.S., 'On the Subjectless "Pseudo-passive" in Standard Dutch and the Semantics of Background Agents', C. Li (ed.), *Subject and Topic*, New York, 1976, pag. 385-416.
- Klein, S., 'Control of Style with Generative Grammar', *Language* (41), 1965, pag. 619-631.
- Kochan, D.C. (Hrsg.), *Stilistik und Soziolinguistik*, Berlin, 1971.
- 33 + Koenen-Endepols, *Verklarend Handwoordenboek der Nederlandse Taal* Groningen, 1977 (zevenentwintigste druk).
- Kolkhuis Tanke, J.A., *Creatieve zakenbrieven*, Groningen, 1975 (vierde druk).
- 2 Korsten, A.F.A., 'Leesbaarheid van overheidsbrochures en sociale ongelijkheid', *Massacommunicatie* (4), 1976, pag. 90-113.
- 1 Kousbroek, R., *Raymond Queneau, Stijloefeningen*, Amsterdam, 1978.
- 110 Kraak, A., Klooster, W.G., *Syntaxis*, Culemborg, 1968.
- Krallmann, D., *Statistische Methoden in der Stilistische Textanalyse, Ein Beitrag zur Informationserschliessung mithilfe elektronischer Rechenmaschinen*, Bonn, 1966.
- 33 + Kramers' *Nederlands Woordenboek*, Den Haag, 1977 (negentiende druk).

- Kraus, J., Polák, J., 'Text Factors and Characteristics', *Prague Studies in Mathematical Linguistics* (2), 1967, pag. 155-171.
- 18 Kroeber, K., 'Perils of Quantification: the Exemplary Case of Jane Austen's Emma', L. Doležal and R.W. Bailey (eds.), *Statistics and Style*, New York, 1969, pag. 197-213.
- 8 Kusters, T.; 'Het onderzoek naar register', *Reader Register*, interne publikatie vakgroep Nederlands, Leiden, 1977, pag. 203-220.
- Langner, H., 'Sprachschichten und soziale Schichten', *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* (27), 1974, pag. 93-104.
- 30, 39 + 40, 41 Lans, C., *Een nieuwe ambtenaar, een nieuwe stijl*, 's-Gravenhage, 1943.
- Leginski, W., Izzett, R.R., 'Linguistic Styles as Indices for Interpersonal Distance', *The Journal of Social Psychology* (91), 1973, pag. 291-304.
- 1 Lennep, J. van, *Proeve eener verduitsching der Grondwet*, Amsterdam, 1844.
- Levin, S.R., 'Deviation - Statistical and Determinate - in Poetic Language', *Lingua* (12), 1963, pag. 276-290.
- 25 Lindner, J.J., 'Afstotelijke taal politici opent wereld van onbegrip', *De Volkskrant*, 1 november 1975.
- Love, G.A., Payne M. (eds.), *Contemporary Essays on Style, Rhetoric Linguistics, and Criticism*, Glenview, Ill., 1969.
- 1 Lulofs, B.H., *Vlugtige woorden over Nederlandsche taalzuivering en taalverrijking, inzonderheid ook met opzigt tot hetgene de nieuwe wetboeken daartoe kunnen en zullen bijdragen*, Groningen, 1826.
- Martin, W., *De inhoud van krant en roman; een frequentieonderzoek*, Antwerpen, z.j. (1968).
- Martin, W., *Analyse van een vocabularium met behulp van een computer*, Brussel, 1970.
- Martin, W., 'Het gebruik van de computer bij taalonderzoek', *Levende Talen*, 1971, pag. 41-53.
- Martin, W., 'Taal en tal. Enkele statistische aspecten van kwantitatieve taalkunde', *Intermediair* (11, nr. 40), 1975, pag. 39-47.
- Martin, W., 'Taal en tal', *Handelingen van het vierendertigste Nederlands Filologencongres*, Amsterdam, 1976, pag. 179-197.
- Martin, W., 'Preliminaire opmerkingen over kwantitatieve lexicologie', P.G.J. van Sterkenburg (red.), *Lexicologie. Een bundel opstellen voor F. de Tollenaere*, Groningen, 1977, pag. 189-196.
- Messing, G.M., 'The Impact of Transformational Grammar upon Stylistics and Literary Analysis', *Linguistics* (66), 1971, pag. 56-73.
- 37, 42 + Meijers, J.A., *Mens en taal*, Amsterdam, 1965.
- Michel, G., 'Stil und Expressivität', *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* (27), 1974, pag. 132-140.
- 8, 10, 17, 19 Milic, L.T., *A Quantitative Approach to the Style of Jonathan Swift*, The Hague, 1967.
- Milic, L.T., 'Autocoding in Computational Stylistics', B.B. Kachru and H.F.W. Stahlke (eds.), *Current Trends in Stylistics*, Edmonton, 1972, pag. 263-273.

- 43 + Milot, C., 'Enig kwaad gesputter nopens ambtelijke stijl in de troonrede', *Contact* (Personeelsblad van het Ministerie van Economische Zaken), 1953 (oktober).
- 39 + Milot, C., 'Van-de-voor-de-variété', *Contact* (Personeelsblad van het Ministerie van Economische Zaken), 1957 (april).
- Moerk, E.L., 'Quantitative Analysis of Writing Styles', *Journal of Linguistics* (6), 1970, pag. 223-230.
- 38, 41 + Moerdijk, J.M., Sluimer, P., *Taalgids voor de ambtenaar*, 's-Gravenhage, 1976 (zesde druk). *MAG 105 F 330*
- 42, 43 Nivette, J., 'Het formaliseren van de syntaxis, frequentie-onderzoek van 10.000 Nederlandse tagmemen met behulp van een computer', *Handelingen van het XXXe Vlaams Filologencongres*, Brussel, 1977, pag. 246-252.
- Ohmann, R., 'Generative Grammars and the Concept of Literary Style', *Word* (20), 1964, pag. 423-439.
- Oksaar, E., 'Stilstatistik und Textanalyse', H. Backes (Hrsg.), *Festschrift für Hans Eggers zum 65. Geburtstag*, Tübingen, 1972, pag. 630-648.
- 1 Orwell, G., 'Politics and the English Language', *The Collected Essays, Journalism and Letters of George Orwell* (vol. IV), London, 1968, pag. 127-140.
- Pauwels, J.L., 'Statistisch onderzoek van de Nederlandse zinsbouw', *De Nieuwe Taalgids* (Van-Haeringennummer), 1970, pag. 93-100.
- 26, 38, 39, 42 + Pieete, A., *Incorrect Nederlands*, Purmerend, 1973 (zesde druk).
- Plett, H.F., 'Concepts of Style: a Classificatory Approach', *Language and Style* (12), 1979, pag. 268-281.
- 40 + Poelje, G.A. van, 'Uit de woordensmidse van wetgeving en bestuur', *Bestuurswetenschappen* (21), 1957, 205-215.
- 31 + Poelje, G.A. van, *Persoonlijke onpersoonlijkheid. Opmerkingen over het ambtelijk taalgebruik* (rede jaarvergadering Maatschappij Nederlandse Letterkunde), Alphen aan den Rijn, 1960.
- 110 Pollmann, T., *Oorzaak en handelende persoon. De beschrijving van passieve zinnen in de Nederlandse grammatica*, Haarlem, 1975.
- Posner, R., 'The Use and Abuse of Stylistic Statistics', *Archivum Linguisticum* (15), Glasgow, 1963, pag. 111-139.
- 68 Pijnenburg, W., 'Enige kwantitatieve aspecten van het Nederlands en het Middelnederlands', *Glott* (2), 1979, pag. 23-29.
- 2 Denkema, J., 'Duidelijke taal. Problemen bij het vereenvoudigen van overheidsteksten', *Intermediair* (12, nr. 38), 1976, pag. 7-13. *Schiedam*
- Revzin, I.O., 'Generative Grammars, Stylistics and Poetics', A.J. Greimas, R. Jakobson e.a., *Sign, Language, Culture*, The Hague, 1970, pag. 558-569.
- 12 Riffaterre, M., 'Stylistic Context', *Word* (16), 1960, pag. 207-218.
- Riffaterre, M., 'The Stylistic Function', H.G. Lunt (ed.), *Proceedings of the Ninth International Congress of Linguistics*, The Hague, 1964, pag. 316-323.
- 30, 35 + Ritter jr., P.H., 'De ambtelijke taal', J.H. de Goede jr. (red.), *Het orkest der overheid. Ambtenaren op hun post*, Alphen aan den Rijn, 1939, pag. 99-125.
- 37, 40 Rogier, L.J., Pieete, A., *Nederlandse Briefstijl*, Stijlcritische inleiding tot de Nederlandse correspondentie voor het economisch onderwijs en voor zelfstudie, Zwolle, 1954 (vierde druk).

- Rosengren, J., 'The Quantitative Concept of Language and its Relation to the Structures of Frequency Dictionaries', *Etudes de linguistique appliquée*, serie nr. 1, 1971, pag. 103-126.
- 25 + Run, H. van, 'Wie dit leest is gek; wartaal en wantaal in nota's en rapporten', *Jeugd en Samenleving* (6), 1976, pag. 3-12.
- Russell, W.M., 'Linguistic Stylistics', *Linguistics* (65), 1971, pag. 75-82.
- 7, 12, 14, 18 Sandell, R., *Linguistic Style and Persuasion*, London, 1977.
- 8, 12 Sanders, W., *Linguistische Stilistik, Grundsätze der Stilanalyse sprachlicher Kommunikation*, Göttingen, 1977.
- Sankoff, D. (ed.), *Linguistic Variation; Models and Methods*, New York, 1978.
- 12 Saporta, S., 'The Application of Linguistics to the Study of Poetic Language', T.A. Sebeok (ed.), *Style in Language*, Cambridge Mass., 1975, pag. 82-93 (zesde druk).
- Schmidt, F., 'Satz und Stil', H. Kreuzer en R. Gunzenhäuser (Hrsg.), *Mathematik und Dichtung*, München, 1971, pag. 159-171 (vierde druk).
- 160 Sedelow, S.Y., Sedelow jr., W.A., 'Models, Computing and Stylistics', B.B. Kachru and H.F.W. Stahlke (eds.), *Current Trends in Stylistics*, Edmonton, 1972, pag. 275-286.
- Seiffert, H., *Stil heute. Eine Einführung in die Stilistik*, München, 1977.
- Siertsema, B., 'Review of A. Ellegård, H. Karlgren en H. Spang-Hansen, Structures and Quanta, Three Essays on Linguistic Description', *Lingua* (20), 1968, pag. 433-435.
- 80 Spang-Hansen, H., 'Sentence Length and Statistical Linguistics', *Structures and Quanta, Three Essays on Linguistic Description, Dedicated to Paul Diderichsen*, Copenhagen, 1963, pag. 58-72.
- 7, 17, 18 Spencer, J., Gregory, M., 'An Approach to the Study of Style', J. Spencer (ed.), *Linguistics and Style*, London, 1964, pag. 57-105.
- Spiewok, W., 'Funktionale Grammatik und Stiltheorie', *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* (27), 1974, pag. 223-230.
- 9, 13-16 Spillner, B., *Linguistik und Literaturwissenschaft. Stilforschung, Rhetorik, Textlinguistik*, Stuttgart, 1974.
- Stutterheim, C.F.P., 'Modern Stylistics', *Lingua* (1), 1948, pag. 410-426 en (3), 1952, pag. 52-63.
- 26 + *Taal van Den Haag*. Voordrachten gehouden tijdens het tiende congres van het Genootschap Onze Taal, *Onze Taal* (45), 1976, pag. 1-30.
- 34, 36, 37 + *Taalbeheersing in de Administratie*, Kortrijk-Heule, 1962-1967.
38, 41, 44
- Tallentire, D.R., 'Mathematical Modelling in Stylistics: its Extent and General Limitations', R.A. Wisbey (ed.), *The Computer in Literary and Linguistic Research*, Cambridge, 1971, pag. 117-129.
- Thorne, J.P., 'Stylistics and Generative Grammars', *Journal of Linguistics* (1), 1965, pag. 49-59.
- Thorne, J.P., 'Generative Grammar and Stylistic Analysis', J. Lyons (ed.), *New Horizons in Linguistics*, Harmondsworth, 1970, pag. 185-197.
- Tollenaere, F. de, 'Coling 76 (International Conference on Computational Linguistics)', *Wetenschappelijke Tijdingen* (36), 1977, pag. 57-64.

- Turner, G.W., *Stylistics*, Harmondsworth, 1973.
- Uit den Boogaart, P.C., 'Sampling van tekstfragmenten uit Nederlandse dagbladen', *ITL-Review for Applied Linguistics* (10), 1970, pag. 25-33.
- 4, 13, 47-78, 80 Uit den Boogaart, P.C. (red.), *Woordfrequenties in geschreven en gesproken Nederlands*, Utrecht, 1975.
- 26 Ule, C.H., 'Die Sprache in der Verwaltung', *Muttersprache* (70), 1960, pag. 363-373.
- 16, 19-21, 142 Ullmann, S., *Language and Style*, Oxford, 1964.
- Ullmann, S., *Meaning and Style*, Oxford, 1973.
- 43 + *Vademecum Ministerie van Onderwijs en Wetenschappen*, 's-Gravenhage, 1976.
- Valge, J., 'A Possibility of Describing Lexical Systems', *Journal of Linguistic Calculus* (2), 1977, pag. 44-59.
- 34 + Veering, J., 'De ingdood', *Onze Taal* (32), 1963a, pag. 37-39.
- 38 + Veering, J., 'Spanning in de zin', *Onze Taal* (32), 1963b, pag. 63-64.
- 28 + Veering, J., *Mogelijkheden en moeilijkheden van taalverzorging*, Delft, 1966.
- 34 + Veering, J., 'Het stervende werkwoord', *Onze Taal* (38), 1969, pag. 53-59.
- 36 + Veering, J., 'Het taalgebruik in een rapport', H. de Boer e.a., *Schriftelijk Rapporteren*, Utrecht/Antwerpen, 1971, pag. 105-144 (achtste druk).
- 28, 29, 38 + Veering, J., 'Knelpunten van procedurele aard', *Onze Taal* (43), 1974, pag. 53-59.
- 31 + Veld, J. in 't, 'Prof. dr. G.A. van Poelje als voortrekker en wegbereider', *Bestuurswetenschappen* (31), 1977, pag. 6-14.
- Verriijn-Stuart, A.A., 'De computertaal in de taal- en letterkunde', *Forum der Letteren* (16), 1975, pag. 200-219.
- 85 *Verslag van de Commissie Duidelijke Taal*, intern commissierapport, Den Haag, 1974.
- 47, 80, 85, 86 *Verslag Werkgroep Frequentie-Onderzoek van het Nederlands*. Een onderzoek naar kwantitatieve, lexicale en grammaticale verschijnselen in het Nederlands, Eindhoven, 1974 (niet gepubliceerd).
- 91, 145, 146
- 68 Vervoorn, A.J., 'Woordfrequenties: anatomie van onze woordenschat', *Onze Taal* (44), 1975, pag. 53-60.
- 24, 85 + Vervoorn, A.J., 'Jacob van Lennep en de taal der Grondwet', *Spiegel Historiael* (13), 1978, pag. 638-645.
- 1 Vierssen Trip, G.W. van, *De Gelaarsde Kat*, vertelling van een merkwaardige strafzaak, Den Haag, 1969.
- 2 Vondeling, A., Renkema, J., *De troonrede van Ridderzaal naar huiskamer*, 's-Gravenhage, 1976.
- 47-78 *Voorschriften lexicale codering in het systeem C₃C*. Werkgroep Frequentie-Onderzoek van het Nederlands, Eindhoven, 1974 (niet gepubliceerd).
- 115, 150
- 41 + Vooy's, C.G.N. de, 'Is dat goed Nederlands?', *De Nieuwe Taalgids* (35), 1941, pag. 36-40.
- Vriendt-de Man, M.J. de, *Frequentie van woorden en structuren in spontaan gesproken Nederlands*, Brussel, 1971.
- 1 Vries, C.W. de, J.R. Thorbecke, *Bijdrage tot de herziening der Grondwet*, 's-Gravenhage, 1948.
- 26, 80 Wagner, H., *Die deutsche Verwaltungssprache der Gegenwart*, Düsseldorf, 1970.

- Wells, R., 'Nominal and Verbal Style', T.A. Sebeok (ed.), *Style in Language*, Cambridge Mass., 1975, pag. 213-221 (zesde druk).
- Williams, C.B., 'A Note on the Statistical Analysis of Sentence Length as a Criterion of Literary Style', L. Doležel and R.W. Bailey (eds.), *Statistics and Style*, New York, 1969, pag. 69-75.
- Wildgen, W., *Kommunikativer Stil und Sozialization. Ergebnisse einer empirischen Untersuchung*, Tübingen, 1977.
- Winter, W., 'Relative Häufigkeit syntaktischer Erscheinungen als Mittel zur Abgrenzung von Stilarten', *Phonetica* (7), 1961, pag. 193-216.
- Winter, W., 'Styles as Dialects', L. Doležel and R.W. Bailey (eds.), *Statistics and Style*, New York, 1969, pag. 3-9.
- 32 + Woordenboek der Nederlandsche Taal, deel VII, 's-Gravenhage, 1926.
Woordfrequenties, zie Uit den Boogaart (1975).
- 31, 40 + Wijngaarden, H.R., 'Gezag en Gezagserkenning', *Macht over Macht*, Baarn, 1975, pag. 41-54.
- 80 Yule, G.U., 'On Sentence-Length as a Statistical Characteristic of Style in Prose', *Biometrika* (30), 1939, pag. 363-390.
- Yule, G.U., *The Statistical Study of Literary Vocabulary*, Cambridge, 1944.

REGISTER

a

Aanhangsel 81, 88, 89, 92
 aanvoegende wijs 43, 57, 129
 aanwijzende bijwoorden 62, 69, 72
 abstracte woorden 161
 achterzetsels 64
 adjectiva-nomina ratio 161
 adjectieven 52, 128, 156
 - aantal verschillende 156
 - adverbiaal gebruikte 53
 - substantivisch gebruikte 52, 105, 148
 adjectivisch gebruikte substantieven 51
 afkortingen 40, 127
 algemene oordelen 27, 30, 144
 ambtelijk taalgebruik 1, 24, 45, 84
 archaïsmen 42, 132, 157

b

buitentalige factoren 6, 7
 Bijlagen 81, 88, 89
 bijvoeglijke naamwoorden 41
 bijvoeglijke bijzinnen 144
 bijwoorden 41, 61, 128
 - aantal verschillende 156
 - aanwijzende en onbepaalde 62, 69, 72
 - betrekkelijke 62
 - betrekkelijke voornaamwoordelijke 63, 119
 - voornaamwoordelijke 62, 64, 121, 154, 161
 - vragende 62
 - vragende voornaamwoordelijke 63
 bijzinnen 38, 39, 118, 154

c

cardinalia 60
 clichématig taalgebruik 43, 129
 codeermogelijkheden, verschillende 62, 63, 65, 66, 74
 codeersysteem C₃C 5, 47, 48, 104, 143, 160
 - nauwkeurigheid 100
 - nomenclatuur 69
 - niet gemaakte onderscheidingen 70
 - overzicht 170
 - slordigheden 73
 - systeemdwang 72
 - vierde cijfer 68, 99
 - voorbeelden 50-68
 Commissie Duidelijke Taal 1, 24, 26, 83, 85

conjuncties 65
 conjunctieven 57
 corpora 104
 correspondentie regering-Statens-Generaal 79, 81, 83, 84, 93

d

dat 37
 - als relativum 154
 - als voegwoord 153
 deelwoorden 56, 71, 106, 161
 - substantivisch gebruikte 106, 148
 demonstrativa 59
 D, O en W, verschillen 159

e

eerste persoon, verwijzingen naar 36, 113, 152
 eigennamen 51, 117, 143, 153

f

fragmentenselectie 90
 frequentielijsten 20, 23, 75, 76, 100, 129

g

gemiddelde zinslengte 80, 118, 146, 220, 221
 germanismen 43
 gespecificeerde oordelen 27

h

homonymie 20, 62, 63, 65, 66
 hulp- en koppelwerkwoorden 55, 111, 149-151
 hypotheses 103-132
 - en oordelen 141-143
 - indeling van 143
 - tegenstrijdigheid in 144
 - overzicht 137-140

i

imperatieven 57, 117, 142, 143, 153
 indefinita 59, 114, 162
 infinitieven (substantivisch) 55, 105, 148
 '-ing'/'-ingen', woorden op - 34, 108, 141, 148
 interjecties 51, 53, 63, 67, 72, 117, 143, 153
 interrogativa 59
 intransitief gebruikte werkwoorden 54

j

juridisch taalgebruik 24, 45, 84

k

kamerstukken 81, 84, 90
kamervragen 81, 153
kanselarijstijl 32
ketens van voorzetsels en lidwoorden 39, 125, 155
klemconstructies 26, 39, 123, 144, 155
- aantal woorden per - 124
- met deelwoord of infinitief 124, 155
- met 'te' 155
koppelwerkwoord 55, 110, 149
kwalitatieve stilistiek 16, 19, 21, 22
kwantitatieve stilistiek 16, 19, 21, 22

l

leesbaarheid 2, 36
leesbaarheidsonderzoek 36, 127
lijdende vorm 26, 35, 108, 149
- bepalingen met door 109, 151
- criteria 110
- met infinitief 151
- hulpwerkwoorden 111
- met intransitief deelwoord 151
- twijfelgevallen 110
- vergelijking 150

m

'men' 36, 114, 152
moeilijke woorden 40, 127, 156

n

naamwoordstijl 33, 104, 143, 147
nevenschikkende voegwoorden 65, 119, 154
'niet-', 'non-', 'on-', woorden met - 126, 127, 156
nomina-verba ratio 161
norm, verwante 10, 11, 13

o

O, D en W, verschillen 159
omhaal van woorden 40, 41
onbepaalde bijwoorden 62, 69, 72
onderschikkende voegwoorden 65, 119, 154
onderzoeksmateriaal, bezwaren 82
oneigenlijke spaties 98
ontkennende woorden 125, 156
ontkenningen 41
Oog op Tekst 100, 106, 108, 112, 114, 115, 116, 120, 121, 123, 124
oordelen 1, 3, 24-46
- de aard van de 28
- algemene 27, 30, 144
- bevestigde 158
- buiten het C₃C-kader 102

oordelen (vervolg)

- gespecificeerde 27
- niet bevestigde 158
- normatieve 4
- tegenstrijdige 45
- in trefwoorden 32
- vaagheid van de 29
- verzamelen van 24
- de waarde van de 44
- operationalisering 141
- over zinsbouw, woordkeus 29, 33
ordinalia 61

p

paginaselectie 88
passieve vorm 35
personalia 58, 111, 112, 152
persoonsvormen 55, 57, 118, 154
pleonastisch taalgebruik 41, 141
politiek taalgebruik 24, 45, 84
populatie 89, 90, 91, 134, 136, 146
possessiva 58, 111-113, 152
pre- en postposities 63
pronomina 58, 59
- in verouderde naamvalsvormen 130, 131, 157
- zelfstandig en bijvoeglijk gebruikt 72

r

reflexief gebruikte werkwoorden 55
reflexiva 58
register-onderzoek 8
relativa 59, 119, 154
retrogradelijsten 101, 108, 161

s

samengestelde werkwoorden 64, 121, 154
spaties, oneigenlijke 98
standaarddeviatie 118, 196
steekproef 86-92
- arbitraire beslissingen 87
- aselecte 90
- omvang 86
- overzicht 93
- representativiteit 83, 90, 145
stilistiek 6, 7
- buitentalige factoren 6, 7
- kwalitatieve 16, 19, 21, 22
- kwantitatieve 16, 19, 21, 22
- object van 6
stijl 2, 3, 6, 8, 9, 10
- als afwijking van een norm 10, 11, 13
stijltheorie 6
stijlopvattingen 8
subcorpus 104
substantieven 34, 50, 70, 78, 105, 148
- adjectivisch gebruikte 51
- in verouderde naamvalsvormen 130, 157
- de tien frequentste 163

substantivisch gebruikte
 - adjectieven 52, 78, 105, 148
 - deelwoorden 78, 106, 148
 - infinitieven 55, 78, 105, 148

t

taalgebruik

- ambtelijk 1, 24, 32, 45, 84
 - archaisch 42, 129, 157
 - clichématig 43, 129
 - in dagbladen 4, 13, 159
 - ingewikkeld 37, 118, 153
 - juridisch 24, 45, 84
 - in opiniebladen 4, 13, 159
 - onpersoonlijk 33, 104, 147
 - pleonastisch 41, 141
 - politiek 24, 45, 84
 - in populair-wetenschappelijke boeken 4, 13, 85, 159
 - traditioneel 41, 129, 157
 - type - 2, 3
 - variatie in 2, 3
 - wijddopig 40, 128, 156
 taal van 'Den Haag' 24
 tangconstructies 38, 120, 141, 142, 154
 'te' 39, 64, 119, 155
 tegenstrijdige oordelen 45
 tekstbewerking 97
 TO-StG, verschillen 158
 toetsing
 - betrouwbaarheidsinterval 136
 - van hypothesen 133
 - onbetrouwbaarheidsdrempel 133
 - significantie 135
 - overzicht van de resultaten 137-140
 - 2 x 2-tabel 134
 - vergelijking met (deel)totaal 135
 transitief gebruikte werkwoorden 54
 trefwoorden, oordelen 32
 tussenzinnen 38, 118, 154
 type taalgebruik 2, 3
 type-token ratio 161

u

'u' 36, 114, 152

v

vaagheid van oordelen 29
 vakjargon 40
 variatie in taalgebruik 2, 3
 verbogen deelwoorden 56, 78, 106, 161
 verwante norm 10, 11, 13
 verwijzingen naar eerste persoon 36, 113, 152
 vierde cijfer in C_3C 68, 99
 voegwoorden
 - nevenschikkende 65, 119, 154
 - onderschikkende 65, 119, 154
 - van vergelijking 66, 119, 154

voegwoord plus voegwoord 38, 122, 155
 voornaamwoordelijke bijwoorden 62, 64, 121, 154, 161
 voorzetsels 64, 107, 148
 - verbogen 131, 157
 - en lidwoorden, ketens van 39, 125, 155
 voorzetselgroepen 64, 121, 154
 vraag- en uitroepzinnen 36, 115, 152
 vraagzinnen met persoonsvorm 115
 - met vraagwoord 116
 vragende bijwoorden 62
 vragende voornaamwoordelijke bijwoorden 63
 vreemde woorden 40, 127

w

W, D en O, verschillen 159
 waarde van de oordelen 44
 werkwoorden 54, 107, 149
 - gesubstantiveerde 34, 149
 - hulp- en koppel 55, 111, 149
 - intransitief gebruikte 54
 - reflexief gebruikte 55
 - samengestelde 64, 121, 154
 - transitief gebruikte 54
 werkwoordsvormen 54
 wetsteksten 88
 woorden
 - abstracte 161
 - moeilijke 40, 127, 156
 - met 'niet-', 'non-', 'on-' 126, 127, 156
 - omhaal van 40, 41
 - ontkennde 125, 156
 - op '-ing' of '-ingen' 34, 108, 141, 148
 - vreemde 40, 127
 - in zelfnoemfunctie 51
 woordkeus, oordelen 33
 woordlengte 128, 156
 woordvormen 76, 161
 woordvoorraad 161
 'worden' als koppelwerkwoord 110

z

zelfstandige naamwoorden 34, 78
 zinsbouw, oordelen 33
 zinslengte 37, 80, 117, 145, 153
 - gemiddelde 118, 146
 - grafieken 147, 220, 221
 'zijn' als hulp- of koppelwerkwoord 108, 111, 149-151